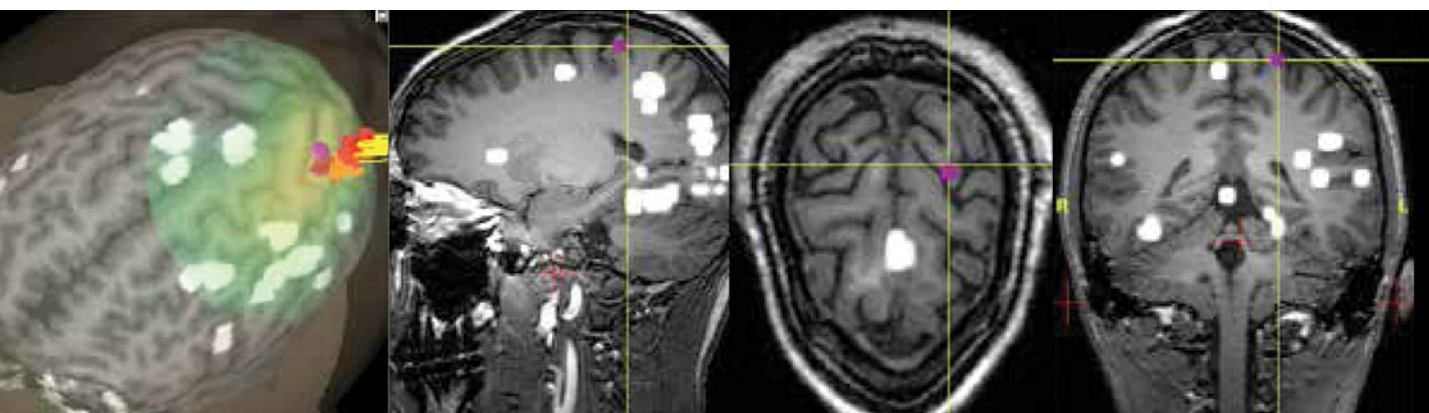


ESSENTIALS OF COGNITIVE NEUROSCIENCE

Bradley R. Postle

WILEY Blackwell

ESSENTIALS OF COGNITIVE NEUROSCIENCE



ESSENTIALS OF COGNITIVE NEUROSCIENCE

Bradley R. Postle

WILEY Blackwell

This edition first published 2015
© 2015 John Wiley & Sons, Ltd.

Registered Office

John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of Bradley R. Postle to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Postle, Bradley R.

Essentials of cognitive neuroscience / Bradley R. Postle.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-46806-7 (pbk.)

1. Cognitive neuroscience--Textbooks. 2. Clinical neuropsychology--Textbooks. I. Title.

QP360.P675 2015

612.8'233--dc23

2014040543

A catalogue record for this book is available from the British Library.

Cover image: Images from an experiment in which repetitive transcranial magnetic stimulation (rTMS) was delivered to targeted regions of the brain while the subject performed a working memory task. White blobs correspond to elevated activity measured in a previous scan, and the purple dot in each image to the rTMS target. These images were acquired during the study published as Postle, Bradley R., Fabio Ferrarelli, Massihullah Hamidi, Eva Ferredoes, Marcello Massimini, Michael Peterson, Andrew Alexander, and Giulio Tononi. 2006. "Repetitive Transcranial Magnetic Stimulation Dissociates Working Memory Manipulation from Retention Functions in Prefrontal, but not Posterior Parietal, Cortex." *Journal of Cognitive Neuroscience* 18: 1712–1722.

Set in 10/12pt Bembo by SPi Publisher Services, Pondicherry, India

BRIEF CONTENTS

Preface **xvi**
Acknowledgments **xix**
Methodology Boxes **xxi**
Walkthrough of Pedagogical Features **xxii**
Companion Website **xxiv**

SECTION I: THE NEUROBIOLOGY OF THINKING 1

- 1 Introduction and History 4
- 2 The Brain 22

SECTION II: SENSATION, PERCEPTION, ATTENTION, AND ACTION 51

- 3 Sensation and Perception of Visual Signals 54
- 4 Audition and Somatosensation 85
- 5 The Visual System 120
- 6 Spatial Cognition and Attention 153
- 7 Skeletomotor Control 196
- 8 Oculomotor Control and the Control of Attention 229

SECTION III: MENTAL REPRESENTATION 259

- 9 Visual Object Recognition and Knowledge 262
- 10 Neural Bases of Memory 286

- 11 Declarative Long-Term Memory 312
- 12 Semantic Long-Term Memory 336
- 13 Short-Term and Working Memory 360

SECTION IV: HIGH-LEVEL COGNITION 395

- 14 Cognitive Control 397
- 15 Decision Making 422
- 16 Social Behavior 449
- 17 Emotion 476
- 18 Language 502
- 19 Consciousness 531

Glossary 562

Index 571

CONTENTS

Preface **xvi**
Acknowledgments **xix**
Methodology Boxes **xxi**
Walkthrough of Pedagogical Features **xxii**
Companion Website **xxiv**

SECTION I: THE NEUROBIOLOGY OF THINKING **1**

1 Introduction and History **4**

KEY THEMES **4**
A BRIEF (AND SELECTIVE) HISTORY **8**
Localization of function vs. mass action **8**
The first scientifically rigorous demonstrations of localization of function **12**
The localization of motor functions **13**
The localization of visual perception **14**
The localization of speech **16**
WHAT IS A BRAIN AND WHAT DOES IT DO? **18**
LOOKING AHEAD TO THE DEVELOPMENT OF COGNITIVE NEUROSCIENCE **19**
END-OF-CHAPTER QUESTIONS **20**
REFERENCES **21**
OTHER SOURCES USED **21**
FURTHER READING **21**

2 The Brain **22**

KEY THEMES **22**
PEP TALK **24**
GROSS ANATOMY **24**
The cerebral cortex **28**
THE NEURON **30**
Electrical and chemical properties of the neuron **31**
How neurons communicate: the action potential **34**
How neurons communicate: the postsynaptic consequences of neurotransmitter release **35**
Different neurotransmitters and different receptors can produce different effects on the postsynaptic cell **37**
Housekeeping in the synaptic cleft **38**
Neuroanatomical techniques exploit the physiology of the neuron **40**
OSCILLATORY FLUCTUATIONS IN THE MEMBRANE POTENTIAL **43**
Neurons are never truly “at rest” **43**
Synchronous oscillation **45**
COMPLICATED, AND COMPLEX **47**
END-OF-CHAPTER QUESTIONS **47**
REFERENCES **48**
OTHER SOURCES USED **48**
FURTHER READING **49**

SECTION II: SENSATION, PERCEPTION, ATTENTION, AND ACTION 51

3 Sensation and Perception of Visual Signals 54

KEY THEMES 54

THE DOMINANT SENSE IN PRIMATES 56

ORGANIZATION OF THE VISUAL SYSTEM 56

The visual field 56

The retinotopic organization of primary visual cortex 58

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – BOTTOM-UP FEATURE DETECTION 63

The V1 neuron as feature detector 63

Columns, hypercolumns, and pinwheels 66

Behind the pinwheel 68

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – INTERACTIVITY 69

Feedforward and feedback projections of V1 69

Corticothalamic loops, center-surround inhibition, and signal-to-noise ratio 72

Setting the state of thalamic neurons: tonic mode vs. burst mode 72

Circularity? It can depend on your perspective 73

The relation between visual processing and the brain's physiological state 73

WHERE DOES SENSATION END? WHERE DOES PERCEPTION BEGIN? 81

END-OF-CHAPTER QUESTIONS 82

REFERENCES 83

OTHER SOURCES USED 83

FURTHER READING 84

4 Audition and Somatosensation 85

KEY THEMES 85

APOLOGIA 87

AUDITION 87

Auditory sensation 87

What is it that we hear? 87

Mechanical transmission and neural transduction 87

Spectral decomposition performed by the basilar membrane 88

Auditory perception 93

The auditory evoked response 93

The organization of auditory cortex 93

Top-down control of auditory transduction 98

Age-related hearing loss 98

Adieu to audition 98

SOMATOSENSATION 99

Transduction of mechanical and thermal energy, and of pain 99

Somatotopy 99

Somatotopy in the rodent: barrel cortex 105

Somatosensory plasticity 112

Use-dependent plasticity 113

The brain mimics the body 113

Mechanisms of somatosensory plasticity 115

Phantom limbs and phantom pain 115

A hypothesized mechanism 115

Helpful or harmful? 116

Proprioception 116

Adieu to sensation 117

END-OF-CHAPTER QUESTIONS 117

REFERENCES 118

OTHER SOURCES USED 118

FURTHER READING 119

5 The Visual System 120

KEY THEMES 120

FAMILIAR PRINCIPLES AND PROCESSES, APPLIED TO HIGHER-LEVEL REPRESENTATIONS 122

TWO PARALLEL PATHWAYS 122

A diversity of projections from V1 122

A functional dissociation of visual perception of *what* an object is vs. *where* it is located 124

Interconnectedness within and between the two pathways 131

THE ORGANIZATION AND FUNCTIONS OF THE VENTRAL VISUAL PROCESSING STREAM 135

Hand cells, face cells, and grandmother cells 135

Electrophysiological recordings in temporal cortex 135

Broader implications of visual properties of temporal cortex neurons 138

A hierarchy of stimulus representation 139

Combination coding, an alternative to the “grandmother cell” 141

A critical role for feedback in the ventral visual processing stream 143

Visual illusions provide evidence for top-down influences 143

Top-down feedback can disambiguate the visual scene 149

TAKING STOCK 149

END-OF-CHAPTER QUESTIONS 150

REFERENCES 151

OTHER SOURCES USED 151

FURTHER READING 152

6 Spatial Cognition and Attention 153

KEY THEMES 153

UNILATERAL NEGLECT:
A FERTILE SOURCE OF MODELS OF SPATIAL COGNITION AND ATTENTION 155

Unilateral neglect: a clinicoanatomical primer 155

Hypotheses arising from clinical observations of neglect 155

THE FUNCTIONAL ANATOMY OF THE DORSAL STREAM 160

Mapping *what* vs. *where* in humans with positron emission tomography 160

Detecting spatial maps with functional magnetic resonance imaging (fMRI) 162

Coordinate transformations to guide action with perception 172

Extracellular electrophysiological studies of monkeys 172

Microstimulation of PPC produces eye movements (and more) 174

FROM PARIETAL SPACE TO MEDIAL-TEMPORAL PLACE 179

Place cells in the hippocampus 179

How does place come to be represented in the hippocampus? 180

From place to long-term memory? 182

THE NEUROPHYSIOLOGY OF SENSORY ATTENTION 182

A day at the circus 183

Attending to locations vs. attending to objects 183

Mechanisms of spatial attention 186

Contradictory findings from fMRI vs. EEG? 186

Effects of attention on neuronal activity 188

Spatial attention 188

Object-based attention 190

TURNING OUR ATTENTION TO THE FUTURE 192

END-OF-CHAPTER QUESTIONS 193

REFERENCES 193

OTHER SOURCES USED 194

FURTHER READING 195

7 Skeletomotor Control 196

KEY THEMES 196

THE ORGANIZATION OF THE MOTOR SYSTEM 198

The anatomy of the motor system 198

The corticospinal tract 198

The cortico-cerebellar circuit 200

The cortico-basal ganglia-thalamic circuits 200

FUNCTIONAL PRINCIPLES OF MOTOR CONTROL 200

The biomechanics of motor control 200

Muscles behave like springs 201

Motor computations in the spinal cord 204

Motor cortex 204

The motor homunculus 205

The neurophysiology of movement 206

EEG and LFP correlates 206

Single unit activity reveals population coding 208

MOTOR CONTROL OUTSIDE OF MOTOR CORTEX 214

Parietal cortex: guiding *how* we move 214

Cerebellum: motor learning, balance, ... and mental representation? 214

Cerebellar circuitry 216

Cerebellar functions 217

Basal ganglia 217
A subcortical gate on cortical activity 218
The role of dopamine 218
 COGNITIVE FUNCTIONS OF THE MOTOR SYSTEM 221
 IT'S ALL ABOUT ACTION 221
 END-OF-CHAPTER QUESTIONS 226
 REFERENCES 226
 OTHER SOURCES USED 227
 FURTHER READING 227

8 Oculomotor Control and the Control of Attention 229

KEY THEMES 229
 ATTENTION AND ACTION 231
 WHYS AND HOWS OF EYE MOVEMENTS 231
 Three categories of eye movements 231
 THE ORGANIZATION OF THE OCULOMOTOR SYSTEM 232
 An overview of the circuitry 232
 The superior colliculus 233
 The posterior system 235
 The frontal eye field 235
 The supplementary eye field 236
 THE CONTROL OF EYE MOVEMENTS, AND OF ATTENTION, IN HUMANS 237
 Human oculomotor control 237
 Human attentional control 238
The endogenous control of attention 241
The exogenous control of attention 242
 THE CONTROL OF ATTENTION VIA THE OCULOMOTOR SYSTEM 243
 Covert attention 243
Covert attention as motor preparation? 243
Empirical evidence for covert attention as motor preparation 243
 Where's the attentional controller? 243
The reentry hypothesis 246
 ARE OCULOMOTOR CONTROL AND ATTENTIONAL CONTROL REALLY THE "SAME THING"? 246

The "method of visual inspection" 250
 "Prioritized Maps of Space in Human Frontoparietal Cortex" 250
 OF LABELS AND MECHANISMS 253
 END-OF-CHAPTER QUESTIONS 255
 REFERENCES 255
 OTHER SOURCES USED 256
 FURTHER READING 257

SECTION III: MENTAL REPRESENTATION 259

9 Visual Object Recognition and Knowledge 262

KEY THEMES 262
 VISUAL AGNOSIA 264
 Apperceptive agnosia 264
 Associative agnosia 265
Prosopagnosia 266
 COMPUTATIONAL MODELS OF VISUAL OBJECT RECOGNITION 266
 Two neuropsychological traditions 266
Cognitive neuropsychology-influenced models 266
Critiques of cognitive neuropsychology-influenced models 268
Parallel distributed processing (PDP)-based models offer a neurally plausible alternative for modeling cognitive processes and architectures 268
 The cognitive neuroscience revolution in visual cognition 271
 CATEGORY SPECIFICITY IN THE VENTRAL STREAM? 271
 Are faces special? 271
"The Guns of August" 271
Questions raised by the Kanwisher et al. (1997) findings 275
 Perceptual expertise 275
 Greeble expertise 276
 Evidence for a high degree of specificity for many categories in ventral occipitotemporal cortex 276

Evidence for highly distributed category representation in ventral occipitotemporal cortex	278
<i>MVPA challenges models of modular category specificity</i>	278
Demonstrating necessity	280
RECONCILING THE IRRECONCILABLE	283
END-OF-CHAPTER QUESTIONS	283
REFERENCES	283
OTHER SOURCES USED	284
FURTHER READING	284

10 Neural Bases of Memory 286

KEY THEMES	286
PLASTICITY, LEARNING, AND MEMORY	288
THE CASE OF H. M.	288
Bilateral medial temporal lobectomy	288
<i>The global amnesic syndrome</i>	288
<i>The organization of memory</i>	290
<i>The subjective experience of amnesia</i>	291
Hippocampus vs. MTL?	291
ASSOCIATION THROUGH SYNAPTIC MODIFICATION	291
The example of Pavlovian conditioning	291
Hebbian plasticity	291
Long-term potentiation	293
<i>Hippocampal circuitry and its electrophysiological investigation</i>	293
<i>The NMDA glutamate receptor as coincidence detector</i>	295
<i>Multiple stages of synaptic strengthening</i>	298
The necessity of NMDA channels for long-term memory formation	298
HOW MIGHT THE HIPPOCAMPUS WORK?	298
Fast-encoding hippocampus vs. slow-encoding cortex	300
Episodic memory for sequences	301
Episodic memory as an evolutionary elaboration of navigational processing	301

WHAT ARE THE COGNITIVE FUNCTIONS OF THE HIPPOCAMPUS?	303
Standard anatomical model	303
Challenges to the standard anatomical model	304
<i>Newer lesioning methods</i>	304
<i>Perceptually challenging behavioral tasks</i>	304
<i>"New roles" for the hippocampus?</i>	305
Consolidation	306
TO CONSOLIDATE	306
END-OF-CHAPTER QUESTIONS	309
REFERENCES	309
OTHER SOURCES USED	310
FURTHER READING	310

11 Declarative Long-Term Memory 312

KEY THEMES	312
THE COGNITIVE NEUROSCIENCE OF LTM	314
ENCODING	314
Neuroimaging the hippocampus	314
<i>Novelty-driven "encoding"</i>	314
<i>Subsequent memory effects in the PFC and MTL</i>	315
<i>Uneven evidence for encoding-related activity in hippocampus</i>	318
Incidental encoding into LTM during a short-term memory task	318
<i>Signal intensity-based analysis of the subsequent memory effect</i>	318
<i>A short-term memory buffer to bridge the temporal gap between to-be-associated events</i>	320
<i>Functional connectivity-based analysis of the subsequent memory effect</i>	320
THE HIPPOCAMPUS IN SPATIAL MEMORY EXPERTS	320
RETRIEVAL	322
Retrieval without awareness	322
Documenting contextual reinstatement in the brain	326

The dynamics of multivariate state changes underlying behavior 326
Implications for “mind reading” ... well, “brain reading” 327
 Familiarity vs. recollection 328
Dual-process models 328
Dual processes of memory retrieval in the rodent 329
 KNOWLEDGE 332
 END-OF-CHAPTER QUESTIONS 332
 REFERENCES 332
 OTHER SOURCES USED 334
 FURTHER READING 334

12 Semantic Long-Term Memory 336

KEY THEMES 336
 KNOWLEDGE IN THE BRAIN 338
 DEFINITIONS AND BASIC FACTS 338
 CATEGORY-SPECIFIC DEFICITS FOLLOWING BRAIN DAMAGE 339
Animacy, or function? 339
A PDP model of modality-specificity 340
The domain-specific knowledge hypothesis 341
 THE NEUROIMAGING OF KNOWLEDGE 342
The meaning, and processing, of words 342
An aside about the role of language in semantics and the study of semantics 344
 PET scanning of object knowledge 344
Knowledge retrieval or lexical access? 347
Knowledge retrieval or mental imagery? 349
 THE PROGRESSIVE LOSS OF KNOWLEDGE 349
Primary Progressive Aphasia or Semantic Dementia, what's in a name? 349
Nonverbal deficits in fluent primary progressive aphasia? 352
The locus of damage in fluent primary progressive aphasia? 353
Distal effects of neurodegeneration 354
Entente cordiale 355
 NUANCE AND CHALLENGES 356

END-OF-CHAPTER QUESTIONS 357
 REFERENCES 357
 OTHER SOURCES USED 358
 FURTHER READING 359

13 Short-Term and Working Memory 360

KEY THEMES 360
 “PROLONGED PERCEPTION” OR “ACTIVATED LTM?” 362
 DEFINITIONS 362
 ELEVATED, SUSTAINED ACTIVITY 365
Early focus on role of PFC in the control of STM 365
Single-unit delay-period activity in PFC and thalamus 365
The implications, and influence, of these first studies of PFC electrophysiology 368
 Working memory 368
The extracellular electrophysiology of “working memory” 369
A third(?) generation of PFC lesion studies 374
STM vs. working memory 375
 A BRAVE NEW WORLD OF MULTIVARIATE DATA ANALYSIS 382
The tradition of univariate analyses 382
 MVPA of fMRI 382
Increased sensitivity reveals delay-period storage in “silent” areas 382
Increased specificity reveals absence of stimulus information in elevated delay-period activity 384
The neural bases of STM 386
 Retrospective MVPA of single-unit extracellular recordings 386
A population-level analysis of “Dynamic Coding for Cognitive Control in Prefrontal Cortex” 386
 THREE-QUARTERS OF A CENTURY 391
 END-OF-CHAPTER QUESTIONS 391
 REFERENCES 391
 OTHER SOURCES USED 393
 FURTHER READING 393

SECTION IV: HIGH-LEVEL COGNITION 395

14 Cognitive Control 397

KEY THEMES 397

THE LATERAL FRONTAL-LOBE SYNDROME 399

Environmental dependency syndrome 400

Perseveration 401

Electrophysiology of the frontal-lobe syndrome 402

Integration? 402

MODELS OF COGNITIVE CONTROL 405

Developmental cognitive neuroscience 405

The A-not-B error 405

Is the A-not-B error a failure of working memory? 405

Generalizing beyond development 406

What makes the PFC special? 408

Properties of the PFC 408

Influence of the DA reward signal on the functions of PFC 409

Combining a reward prediction error signal with a flexible PFC to overcome perseveration 412

NEURAL ACTIVITY RELATING TO COGNITIVE CONTROL 413

Error monitoring 413

The ERN/Ne 413

Performance monitoring and the dorsomedial frontal cortex 413

Oscillatory synchrony 414

Thalamic control of cortical synchrony 415

WHERE IS THE CONTROLLER? 417

END-OF-CHAPTER QUESTIONS 418

REFERENCES 419

OTHER SOURCES USED 420

FURTHER READING 421

15 Decision Making 422

KEY THEMES 422

BETWEEN PERCEPTION AND ACTION 424

PERCEPTUAL DECISION MAKING 424

Judging the direction of motion 424

Microstimulation of MT 424

LIP 426

Perceptual decision making in humans 430

VALUE-BASED DECISION MAKING 433

The influence of expected value on activity in LIP 434

Decision variables emergent from motor plans 435

Common currency in the omPFC 435

Transitivity, value transitivity, and menu invariance 436

Evidence from the nonhuman primate 436

Evidence from humans 439

vmPFC or OFC? 440

Has neuroeconomics taught us anything about the economics of decision making? 440

FORAGING 441

I CAN'T DECIDE! 442

NEXT STOP 442

END-OF-CHAPTER QUESTIONS 446

REFERENCES 446

OTHER SOURCES USED 448

FURTHER READING 448

16 Social Behavior 449

KEY THEMES 449

TRUSTWORTHINESS: A PREAMBLE 451

Delaying gratification: a social influence on a "frontal" class of behaviors 451

THE ROLE OF vmPFC IN THE CONTROL OF SOCIAL COGNITION 452

Phineas Gage 452

Horrible accident 452

Sequelae 453

Contemporary behavioral neurology 453

Linking neurological and psychiatric approaches to social behavior 455

THEORY OF MIND 455

The ToM network 458

Two asides about anterior paracingulate cortex 458

The temporoparietal junction (TPJ) 458
Challenge to the specificity of TPJ for ToM 459
 False beliefs (?) about Rebecca Saxe's mind 461
fMRI with higher spatial resolution 461
rTMS of TPJ during moral judgments 462
Moral judgments, TPJ, and autism spectrum disorder 464
 A final assessment of the role of TPJ in ToM mentalization 464
 OBSERVATIONAL LEARNING 467
 TRUSTWORTHINESS, REVISITED 473
 END-OF-CHAPTER QUESTIONS 473
 REFERENCES 473
 OTHER SOURCES USED 475
 FURTHER READING 475

17 Emotion 476

KEY THEMES 476
 EMOTION PROCESSING AND MENTAL HEALTH 478
 TRUSTWORTHINESS REVISITED – AGAIN 478
 A role for the amygdala in the processing of trustworthiness 478
Neuropsychological evidence 478
fMRI evidence 478
 Implicit information processing by the amygdala 480
 THE AMYGDALA 480
 Klüver–Bucy syndrome 480
 Pavlovian fear conditioning 482
Induction 482
Distinguishing avoidance learning from fear conditioning 484
 Emotional content in declarative memories 485
Pharmacological interference with norepinephrine 485
Emotional declarative memory in patients with amygdala lesions 485
 The amygdala's influence on other brain systems 487
 What emotions does the amygdala process? 488

WHAT IS AN EMOTION? 488
 THE CONTROL OF EMOTIONS 491
 Extinction 491
Research in the rodent 492
The roles of midline frontal cortex 492
 Fear conditioning and extinction in the human 493
 “Real-time,” volitional control of emotions 493
 The good and the bad of emotion regulation 495
 THE AMYGDALA AS AFFECTIVE HUB 495
 END-OF-CHAPTER QUESTIONS 498
 REFERENCES 499
 OTHER SOURCES USED 500
 FURTHER READING 501

18 Language 502

KEY THEMES 502
 A SYSTEM OF REMARKABLE COMPLEXITY 504
 WERNICKE–LICHTHEIM: THE CLASSICAL CORE LANGUAGE NETWORK 504
 The aphasias 504
 The functional relevance of the connectivity of the network 505
 SPEECH PERCEPTION 506
 Segregation of the speech signal 506
The elements of language 506
The speech signal 507
Neural correlates of speech perception 509
 Dual routes for speech processing 509
 GRAMMAR 512
 Genetics 512
 Rules in the brain? 515
 Broca's area 516
The anatomy and connectivity of Broca's area 516
Functional accounts of Broca's area 516
 The electrophysiology of grammar 519
 SPEECH PRODUCTION 520
 A psycholinguistic model of production 521
 Forward models for the control of production 521

Necessary neural substrates?	523
ALL TOGETHER NOW	527
“AROUSING NOVEL IDEAS AND SUBTLE PERCEPTIONS AND JUDGMENTS” ...	
WITH THE WRITTEN WORD	527
END-OF-CHAPTER QUESTIONS	527
REFERENCES	527
OTHER SOURCES USED	529
FURTHER READING	530
19 Consciousness	531
KEY THEMES	531
THE MOST COMPLEX OBJECT IN THE UNIVERSE	533
DIFFERENT APPROACHES TO THE PROBLEM	533
THE PHYSIOLOGY OF CONSCIOUSNESS	534
Neurological syndromes	534
Coma	534
Vegetative state/UWS	535
Recovery of consciousness from coma or UWS	538
Two dimensions along which consciousness can vary	538
Sleep	539
Anesthesia	540
Studying anesthesia with TMS/EEG	540
Studying anesthesia with surface EEG and deep brain EEG	543
Summary across physiological studies	543
BRAIN FUNCTIONS SUPPORTING CONSCIOUS PERCEPTION	545
Are we conscious of activity in early sensory cortex?	546
What intrinsic factors influence conscious visual perception?	547
Bottom-up and top-down influences on conscious visual perception	547
The role of reentrant signals to V1	548
Manipulating extrinsic factors to study conscious vs. unconscious vision	549
THEORIES OF CONSCIOUSNESS	552
Global Workspace Theory	552
Recurrent Processing Theory	553
Integrated Information Theory	555
A practical application of Integrated Information Theory	556
UPDATING THE CONSCIOUSNESS GRAPH	558
END-OF-CHAPTER QUESTIONS	558
REFERENCES	558
OTHER SOURCES USED	560
FURTHER READING	560
Glossary	562
Index	571

PREFACE

Does the world need another cognitive neuroscience textbook? Well, “need” is such a loaded word ... What I’ll try to do here is summarize my motivation for writing this book and, in the process, convey the pedagogical vision behind choices about style, content, and organization.

For over 10 years I have taught cognitive neuroscience to undergraduate students and to graduate students at the University of Wisconsin-Madison. For undergraduates I offer an upper-level seminar that is linked to a large-enrollment lecture course on cognitive psychology. The lecture course is a survey of cognitive psychology, spanning sensation and perception to decision making, language, and consciousness. The upper-level seminar course, the impetus for this book, is populated by 12–15 honors students who are concurrently enrolled in the lecture course. In the seminar, which meets once per week for 75 minutes, we discuss the cognitive neuroscience of what’s been covered in that week’s cognitive psychology lectures. The modal student in this seminar has a background in experimental design and probability and statistics for experimental psychology; only some have any formal training in biology, and fewer still in neuroscience. In this context, I have found that a textbook that attempts to cover “everything that we know” in the field is not effective – it will either be too superficial in its coverage to support substantive discussions in class, or it will be inappropriately lengthy and dense, a cognitive neuroscience analogue of the >1000-page medical-school neuroscience textbook that I lugged around as a graduate student in systems neuroscience. Another approach, which is to assign papers from the primary literature, has proven to be an unsatisfactory alternative, because it results in an inordinate amount

of class time being taken up with the remedial work of filling in background that the assigned papers don’t cover, and explaining methods so that students can fully understand how the experiments were performed and how their results can be interpreted.

Essentials of Cognitive Neuroscience aspires to be a textbook that will support high-level, substantive engagement with the material in settings where contact hours and/or expectations for outside-of-the-classroom preparation time may be constrained.¹ It introduces and explicates key principles and concepts in cognitive neuroscience in such a way that the reader will be equipped to critically evaluate the ever-growing body of findings that the field is generating. For some students this knowledge will be needed for subsequent formal study, and for all readers it will be needed to evaluate and interpret reports about cognitive neuroscience research that make their way daily into the news media and popular culture. The book seeks to do so in a style that will give the student a sense of what it’s like to be a cognitive neuroscientist: when confronted with a problem, how does one proceed? How does one read and interpret research that’s outside of one’s sub-area of specialization? How do two scientists advancing mutually incompatible models interrelate? Most importantly, what does it feel like to partake in the wonder and excitement of this most dynamic and fundamental of sciences?

To translate these aspirations into a book, I have made the following pedagogical and organizational choices:

¹ The companion teaching guide also suggests how the book can serve as the primary text for a comprehensive survey course by supplementing readings from the book with concurrent, complementary readings drawn from the primary literature.

- Background is limited to two preliminary chapters, one on history and key concepts, and one a primer on cellular and circuit-level neurobiology. Specific methods are introduced, in *Methodology Boxes*, throughout the book, at the point at which they are relevant for the research question that is being described. This reflects a deliberate choice not to include a third introductory chapter on “Methods of Cognitive Neuroscience.” In my experience, such chapters are at best pedagogically ineffective, and at worst, counterproductive. As an early-in-the-semester assignment, the details from such a chapter about any particular method can have little significance for the reader, because these details are being presented prior to the substantive questions that motivate the very use of that method. The result is that this material is harder to learn, and less interesting, than it would be if introduced in the context of a scientific question that the reader is learning about. I also believe this format to be inefficient, because the reader invariably ends up needing to return to the Methods chapter to re-read the section on, say, fMRI, or TMS, or optogenetics, when s/he gets into the meat of the book and encounters an experiment that uses one of these methods. Nonetheless, if the instructor or student wants to assign/read all the *Methodology Boxes* all at once, s/he can do so with the help of their listing in a dedicated section of the Table of Contents.
- Important concepts are introduced, and explained, via experiments that illustrate them. These won’t always be the first experiment that is credited with introducing an idea, nor even an experiment performed by the investigator or group that is credited with innovating the idea. Rather, they will be ones that I feel are particularly effective for explicating the concept in question. Further, most of the experiments that make it into this book are considered in a fair amount of depth. In so doing, I seek to model a “best practice” of the cognitive neuroscientist, one that is captured by a quote from the nineteenth-century neuroscientist David Ferrier that is relayed in a *Tip* in *Chapter 16*:

“In investigating reports on diseases and injuries of the brain ... I feel obliged always to go to the fountainhead – dirty and muddy though this frequently turns out.” A corollary to this point is that because cognitive neuroscientists learn about and work with key concepts via empirical findings, almost all of the figures in this book are taken directly from the primary scientific literature, just as they appear in the empirical reports or review papers in which they were published. This is also how I teach, and reflects my conviction that an important skill for students to develop is how to interpret figures as they appear in scientific publications. Thus, the book contains only a minimum of figures/diagrams/cartoons that were developed expressly for pedagogical purposes.
- Because a book targeting “essentials” can’t cover all of the areas of research that fall under the rubric of cognitive neuroscience, each chapter concludes with an annotated list of suggested *Further Reading*. Many of the listed works will provide effective entrées into an area that the book doesn’t cover.
- The companion Web site contains a wealth of information that will contribute importantly to the student’s experience of learning from the book, and to the instructor’s experience of teaching from it. The element about which I’m most excited, and which may be the most unique for a textbook on this topic, is the *Web Clips*. These are 6-to-12-minute narrated audiovisual presentations contributed by over 40 experts in different areas of cognitive neuroscience. From them, students will literally hear from the expert her- or himself about either (a) an important experiment that is summarized in the book and that I think merits a telling from the perspective of s/he who actually performed the study, or (b) an important area of cognitive neuroscience that can’t be covered the book but that the reader really should know about. These *Web Clips*, more so than any other element in the book, generate the frisson of “what it’s like” to work in this field. I am greatly indebted to my colleagues who contributed their time and expertise to this remarkable collection.

- Finally, the book is liberally sprinkled with *Tips*. These are like the asides that one can impart in a lecture – information that “won’t be on the exam,” but may help to enliven or otherwise enrich the material. In some instances they’ll provide background on the origin and/or idiosyncratic use of nomenclature. In others, they’ll comment on the interesting, perhaps unorthodox, career path followed by the scientist whose work is a focus of the main text. In this regard, I was gratified to read the following from a student whose professor taught with a preliminary draft of the book, assigning a written critique of the chapter that the student was to present to the class (in this instance, *Chapter 9*): “I found much of the additional content helped promote understanding ... It is nice to learn more background about influential researchers ... [and] when later discussing the contributions of Nancy Kanwisher and Martha Farah in class, I found these ... bits of background helped me recall details regarding both the research and associated findings for each.”

Although conceived and written for an upper-level undergraduate readership, *Essentials of Cognitive Neuroscience* can also be used for graduate teaching. For example, one associate of mine has used it as the primary text in a general Introduction to Cognitive Neuroscience course in which each week a student presented a chapter to the class and led the ensuing discussion. In my more focused graduate seminar on the Cognitive Neuroscience of Attention and Memory I will assign the relevant chapter(s) as background to supplement each week’s readings from the primary literature. The explicit understanding will be that the book’s content will encompass the foundational knowledge that will be assumed for all of our discussions and assignments.

Those of us who are lucky enough to make a living grappling with some facet of the mind–brain problem have a responsibility to impart to our students a sense of the wonder, awe, and excitement that permeate our professional lives. My hope is that this book and its accompanying features can help some of us meet this responsibility.

ACKNOWLEDGMENTS

The efforts and/or indulgence of many made this book possible. Lab manager Mike Starrett has provided unflagging support at all stages of this project, from tracking down obscure, out-of-print volumes to formatting references to proofreading drafts to filming and editing the author's introduction to each chapter. "*MStarr*: We don't know how to do that ... yet"! Dan Acheson was a wise and resourceful collaborator, and is responsible for much of the content on the book's companion Web page, and for all of the instructional support material. Dan was also a critical guide and sounding board during the writing of *Chapter 18*. Additional colleagues (and one spouse) who also provided valued feedback on portions of the book include Sue Corkin, Olivia Gosseries, Mike Koenigs, Bornali Kundu, Maryellen MacDonald, Jay McClelland, Chris Moore, Pat Mulvey, Tim Rogers, and Giulio Tononi. I am fortunate to work in the company of so many smart and generous individuals. Jeff Johnson taught a graduate seminar at North Dakota State University with first drafts of selected chapters, and feedback from his students and from him was very useful at the revision stage.

The indulgence of graduate students and postdocs in my laboratory who tolerated my engagement with this project was much appreciated, as were the substantive contributions of Bornali Kundu and Adam Riggall.

Colleagues who willingly gave of their time and energy to contribute a *Web Clip* for the companion Web page have added enormously to the pedagogical value of this enterprise, and I am very grateful to each of them: Dan Acheson, Andy Alexander, Chris Baker, Shaowen Bao, Rasmus Birn, Sue Corkin, Clay Curtis,

Mark D'Esposito, Mingzhou Ding, Howard Eichenbaum, Olivia Gosseries, Shawn Green, Peter Hagoort, John-Dylan Haynes, Mike Koenigs, Victor Lamme, John Lisman, Bea Luna, Marcello Massimini, Yuko Munakata, Betsy Murray, Behrad Noodoust, Josef Parvizi, Sean Polyn, Mike Robinson, Tim Rogers, Bas Rokers, Vincenzo Romei, Yuri Saalman, John Serences, Jon Simons, Heleen Slagter, Leah Somerville, Craig Stark, Mark Stokes, Jack Van Horn, Brad Voytek, and Gagan Wig. Wow! What a list!

I'm also grateful to many additional colleagues who took the time to track down and share high-resolution copies of hard-to-find images that now appear as figures in the book.

The team at Wiley, and their associates, have provided much-needed guidance, assistance, and encouragement at every stage of the process. Thanks to Andy for conceiving and proposing this project, and to "Rosie," Karen, Jen, Hazel, and Claire for everything else.

A book such as this is necessarily shaped by one's past experience, and I'm grateful to everyone who has guided and influenced my career in cognitive neuroscience: from pre-graduate school mentors, to professors and classmates and lab mates at MIT, to the team in Area 9 (U. Penn instantiation) and the D'Esposito diaspora, to all the colleagues and students with whom I've interacted in the classroom, in the lab, and in the hallway, since coming to Wisconsin. Professional organizations whose members and programming have enriched my career include the Cognitive Neuroscience Society, the International Society for Behavioral Neuroscience, the Memory Disorders Research Society, the Psychonomic Society, the Society for Human Brain Mapping, and the Society for Neuroscience.

Much of the aforementioned career has been underwritten by generous support from the National Institutes of Health and the National Science Foundation, for which I am grateful. These and other funding entities in the US and around the world, public and private, are necessary and valued contributors to all the knowledge and discovery that is recounted on the pages that follow.

Finally, I am forever indebted to my loving family, without whose support and indulgence this book could not have been written.

Madison, WI

June 2014

METHODOLOGY BOXES

- 1.1 Neuropsychology and the importance of anatomical specificity 15
- 2.1 MRI and DTI – methods for in vivo neuroanatomy 32
- 2.2 Cell biological methods for measuring neuronal connectivity 41
- 2.3 Extracellular recording: action potentials and LFPs 44
- 3.1 Functional tissue imaging 67
- 3.2 Electroencephalography (EEG) 74
- 3.3 Transcranial magnetic stimulation (TMS) 77
- 4.1 Spectral analysis of time-varying signals 91
- 4.2 Optogenetic manipulation of cellular activity 107
- 5.1 The double dissociation of function 126
- 6.1 X-ray computed tomography (CT) and positron emission tomography (PET) 161
- 6.2 Functional magnetic resonance imaging (fMRI) 166
- 8.1 Block designs vs. event-related designs in fMRI (Zarahn, Aguirre, and D'Esposito, 1997) 239
- 8.2 Multivariate pattern analysis (MVPA) of fMRI data 251
- 9.1 Parallel distributed processing (PDP) modeling of cognition 269
- 9.2 Some problems, and solutions, with univariate analyses of fMRI data 274
- 9.3 Electrocorticography (ECoG) 281
- 11.1 Measuring functional connectivity with fMRI 321
- 11.2 Estimating memory signals with an ROC-based procedure 330
- 12.1 Problems in neuropsychology: How definitive is a single case study? A double dissociation? 343
- 12.2 Repetition effects and fMRI adaptation 350
- 13.1 Retrospective MVPA of single-unit extracellular recordings 387
- 14.1 Measuring ERPs in patients with focal cortical lesions 403
- 15.1 A signal detection theory-based framework for perceptual decision making 428
- 16.1 Using transcranial magnetic stimulation (TMS) to alter neural functioning 463
- 16.2 Model-based fMRI 472
- 17.1 Meditation as mental training 496
- 18.1 Voxel-based lesion-symptom mapping (VLSM) of speech production and comprehension 524
- 19.1 Studying sleep (and consciousness) with simultaneous TMS-EEG 541

WALKTHROUGH OF PEDAGOGICAL FEATURES



Key Themes The opening page of each chapter lists the key themes that will be introduced and explored in that chapter, to help the reader navigate the text.



Tips Like the asides that a professor sometimes shares during lecture, these present information that is not fundamental to the content, but that may that help to enliven or otherwise enrich the material.



Methodology Boxes These appear throughout the book, at the point at which an understanding of a particular method is relevant for the research question that is being described.



Thought Boxes These explore in depth a question or problem that is raised in the main body of the text, and encourage deeper critical thinking.

particular direction. That is, one neuron might respond preferentially to a boundary oriented vertically, another to a boundary tilted 30° to the right of vertical, another to a boundary oriented horizontally, and so on, such that, with enough recording from enough neurons, they collected evidence that the full 180° of possible orientations would be covered. (Note that it is only necessary to talk about 180° of possible orientations for a bar because, for example, stimuli of 0° and 120° from the same starting point would be the same stimulus.) The property of firing action potentials most reliably in response to a certain range of orientations has come to be known as "orientation selectivity" and is a hallmark property of many

These provide

the brain, and the same is true for other tissues. In this chapter, we will review the **bioenergetics** of central nervous system – to consider how this knowledge has been acquired. Let's start by considering that no matter how long, or multiply branched, a neuron is, it's part of one cell. Thus, there needs to be a mechanism, for example, for the transport of proteins that are synthesized from tRNA in the nucleus of the cell, into various types of sub-membranous organelles, such as mitochondria, and then to the cell periphery, for example, for the synthesis of neurotransmitters. The **cytoskeleton** is critical for the healthy functioning of the neuron. The reason for bringing this up now is that, for a neuron (whether they have known it or not), an axon has some obvious advantage of **axonal transport** of cytoskeleton in the brain (Molecular Biology 2.2; Figure 2.5).

Real-World Applications Boxes

END-OF-CHAPTER QUESTIONS

3. In what ways does function follow form structure in the visual context?
4. Design an experiment that, if your hypothesis is correct, yields positive evidence for a double-blind design. What is the purpose of a double-blind? How can you support evidence that a single-blind design?
5. Define the functions of the *what* and *where* pathways. Summarize at least three types of evidence for this organizing principle of the visual system.
6. Summarize the types of image processing that occurs at each level of the ventral-stream hierarchy.
7. Define the concept of the "grandmother cell." For what reason is it a poor approximation of human memory, perception or thought, is it preferable?
8. What is a plausible alternative to a "grandmother cell" model of recognition? Summarize the evidence for this.
9. How does perceptual learning affect other objects in the visual space?
10. Define objects based on, image-based models of objects, objects in memory. Name at least one, and explain its role.
11. How does feedback from temporal cortex affect visual perception?

Further Reading

An annotated list of readings

COMPANION WEBSITE

www.wiley.com/go/postle/cognitiveneuroscience

This book has a comprehensive companion website that features a number of useful resources for instructors and students alike.

FOR INSTRUCTORS

Sample syllabi

Exam questions

Web clips

Brief presentations (6 to 12-or-so minutes long) by an expert in cognitive neuroscience or an allied discipline that further elucidate or elaborate on content presented in the text.

Please note that the resources for instructors are password protected and can only be accessed by instructors who register with the site.

FOR STUDENTS

Author-narrated summary videos

Glossary

Interactive multiple-choice questions

Web links

A curated collection of Web sites and video clips (many from YouTube), freely accessible on the internet, whose content complements content presented in the text. Web Link 1.1 deserves a special note because it accesses a 3D, rotatable, “dissectible” human brain with labels and descriptions, developed and hosted by Genes to Cognition Online.

SECTION I
THE
NEUROBIOLOGY
OF THINKING

INTRODUCTION TO SECTION I THE NEUROBIOLOGY OF THINKING

Anyone who can read these words, including *you*, experiences conscious awareness. *What explains this? Is the conscious experience of other people “the same” as my own? Is our conscious experience qualitatively different from what may or may not be experienced by nonhuman animals?* These and similar questions have fascinated some of humankind’s most celebrated thinkers for as long as humans have been leaving records of their cogitation. They are among the most profound that a human being can ask. Not surprising, then, is the fact that questions that relate either directly or indirectly to the phenomenon consciousness and the related construct of *cognition* (i.e., thinking) have been taken up by many different scholarly disciplines, including philosophy, evolutionary biology, anthropology, economics, linguistics, computer science, and psychology. What distinguishes **cognitive neuroscience**, the focus of this book, from these and other disciplines, is its grounding in the methods and traditions of neuroscience, and the primacy that it places on understanding the biological bases of mental phenomena. Now it bears noting that cognitive neuroscience doesn’t concern itself only with consciousness and conscious awareness. Indeed, I feel confident asserting that the vast majority of articles that have been published in, say, the *Journal of Cognitive Neuroscience* (one of many scientific journals publishing peer-reviewed reports of

research in the field), don’t even explicitly address the idea of consciousness, much less use the word. Nonetheless, it is also true that it can be difficult to entertain detailed thoughts about human behavior and the principles that govern it without straying into ideas such as those that opened this paragraph. This clear relationship to profound philosophical questions is one of the qualities that differentiates cognitive neuroscience from other physical and biological sciences, including many other domains of neuroscience.

Like other physical and biological scientists, the cognitive neuroscientist designs and carries out rigorously controlled experiments that yield objective, measurable data, and seeks to relate these data to mechanistic models of how a natural system works. And, as suggested above, the system studied by cognitive neuroscientists is one that is also of direct interest to scholars who study it from a very different perspective (e.g., from philosophy, from anthropology, and from **cognitive psychology**). This overlap turns out to be both a blessing and a curse. It’s a blessing in that the cognitive neuroscientist can draw on ideas and observations from a vast array of rich intellectual traditions. And because we’re studying questions that are fundamental to the human condition, almost any thinking person that a cognitive neuroscientist encounters is likely to be interested in what s/he does for a living.

(Much easier, in the humble opinion of this author, to be a cognitive neuroscientist attending a social function full of strangers than, say, an economist, or a cell biologist, or a particle physicist ...) The curse is found on the flip side of this coin. Because scholars in different fields are often interested in the same object of study (whether it be visual perception, or aesthetics, or antisocial behavior), cognitive neuroscience research can often be characterized by people working in other fields as asking the wrong question, or asking it in the wrong way, or of generating findings that are fundamentally irrelevant for understanding the question at hand. Being aware of this sociological context surrounding cognitive neuroscience is just one piece of knowledge that can be helpful in evaluating the implications and importance of any particular set of facts and ideas that will be described in this book.

COGNITIVE NEUROSCIENCE? OR “HUMAN NEUROSCIENCE”? OR “NEUROSCIENCE-WITH- DIRECT-IMPLICATIONS-FOR- UNDERSTANDING-HUMAN- BEHAVIOR”?

How were the topics to be covered in this book selected? More importantly, when teaching cognitive neuroscience, where does one draw the line that defines the boundaries of the discipline? This is a difficult question that doesn't have a definitive answer. It's really just an accident of history that the confluence of neuroscientific methods and studies of human behavior happened first, or, at least, influentially, with domains of behavior and function that are studied by cognitive psychology (e.g., visual perception, language function, memory), as opposed to, for example, social behavior, personality, emotion, or psychopathology. As a result, one sees the label “cognitive neuroscience” appearing earlier in the literature than, for example, **affective neuroscience** or, certainly, **social cognitive neuroscience**. A consequence of this is that the term “cognitive neuroscience” has come to be used in many contexts, and not always with precisely the same meaning. In one, “cognitive neuroscience” can be used to refer to the tools and methods used to study the neural bases of human behavior (e.g., **brain** scans, electrical recording, brain stimulation). This might seem like a misnomer, however, if the behavior being

studied doesn't directly relate to cognition (e.g., the study of sleep, or the study of depression). The label “cognitive neuroscience” can also seem like a bad fit for research that spans multiple traditional categories of brain and/or behavior. To pick one real-world example, let's consider the study of how neural systems measured with brain scans relate to performance on an economic decision-making task by high trait-anxiety vs. low trait-anxiety prisoners (all classified as “psychopaths”), and the comparison of these data to data from neurological patients with frank damage to the ventromedial prefrontal cortex. Should this be classified as cognitive neuroscience research? On one level, I hope that the answer is “yes,” because we will be considering this research in *Chapter 16*! However, to do so is to not explicitly acknowledge the equally important contributions to this research from clinical psychology, from affective neuroscience, from neuroeconomics, and from neuropsychology.

In view of the above, to capture the interdisciplinary breadth of much of what we'll be considering in this book, might it perhaps have been better to entitle it “Essentials of *Human* Neuroscience”? Here the answer is an unequivocal “no,” because to do so would be to exclude the fact that understanding the neural bases of almost all domains of human behavior requires a thorough knowledge of analogous neural functioning in nonhuman animals. This is because, as we shall see in almost every chapter of this book, technical and ethical limitations of what we can measure in the human require us to draw heavily on the results of research performed with nonhuman animals. Hence the idea, implied in the title of this section, that the most precise title for this book might be “Essentials of Neuroscience with Direct Implications for Understanding Human Behavior.” Now I'm no expert in academic publishing, but I'm fairly confident that, were I to have actually proposed this title, my editor would have rejected it. If nothing else, there probably aren't many universities where one can find a course that's called “Introduction to Neuroscience with Direct Implications for Understanding Human Behavior.” Nonetheless, this is probably as good a summary, in 10 words or less, of what this book hopes to cover.

And so, with these considerations in mind, we'll stick with the label “cognitive neuroscience.” It's not perfect, but if one is comfortable with a reasonably broad definition of **cognition** as thinking, behaving, and the factors on which these depend, then this label will serve us reasonably well.

CHAPTER 1

INTRODUCTION AND HISTORY

KEY THEMES

- Although the phenomenon consciousness and the related construct of *cognition* (i.e., thinking) are the focus of many different scholarly disciplines, what distinguishes cognitive neuroscience is its grounding in the methods and traditions of neuroscience, and the primacy that it places on understanding *the neurobiological bases* of mental phenomena.
- There are two levels at which the term “cognitive neuroscience” is used: broadly, it has come to refer to the neuroscientific study of most domains of human behavior; narrowly, it refers to the study of neural bases of *thinking* – what influences it, what it consists of, and how it is controlled.
- The roots of cognitive neuroscience can be traced back to a nineteenth-century debate over two ways of thinking about brain function that both remain relevant today: localization of function vs. mass action.
- Mid-to-late nineteenth-century research established the validity of localization for three functions: motor control (localized to frontal lobes); vision (localized to occipital lobes); speech production (localized to the left posterior inferior frontal gyrus).
- The motor control research introduced the principle of *topographic representation*, that adjacent parts of the body can be represented on adjacent parts of the cerebral cortex.
- Studying an aspect of cognition requires careful thought about the validity of the function to be studied; and not all aspects of human behavior are equally amenable to cognitive neuroscience research.
- The discipline of cognitive neuroscience could not exist without discoveries yielded by research with nonhuman animals.
- At the dawn of the twentieth century, scientists were studying the brain and behavior from three related, but distinct, perspectives that would eventually give rise to cognitive neuroscience as we know it today: systems neuroscience; behavioral neuroneurology/neuropsychology; and experimental psychology.

CONTENTS

KEY THEMES

A BRIEF (AND SELECTIVE) HISTORY

Localization of function vs. mass action

The first scientifically rigorous demonstrations
of localization of function

The localization of motor functions

The localization of visual perception

The localization of speech

WHAT IS A BRAIN AND WHAT
DOES IT DO?

LOOKING AHEAD TO THE DEVELOPMENT
OF COGNITIVE NEUROSCIENCE

END-OF-CHAPTER QUESTIONS

REFERENCES

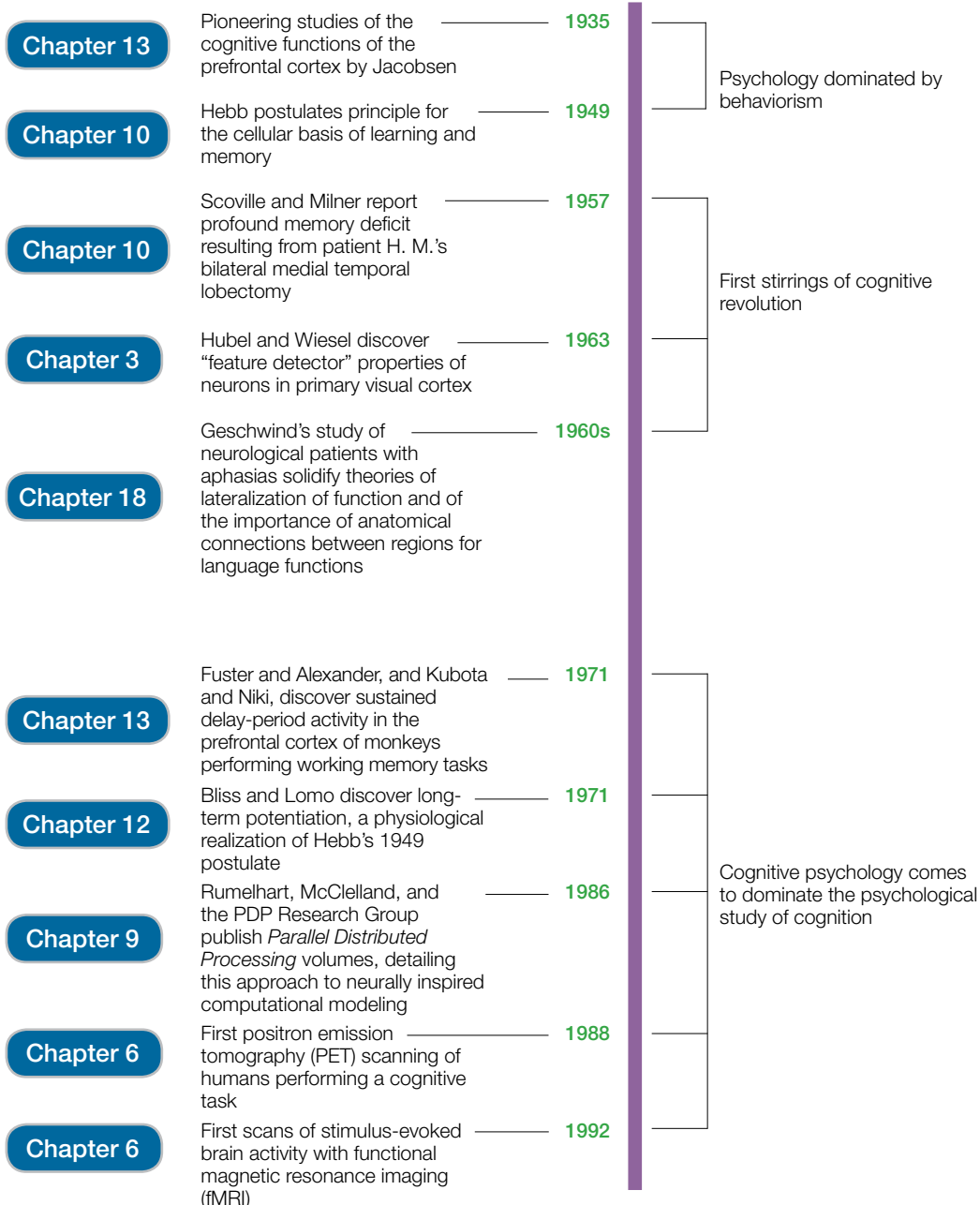
OTHER SOURCES USED

FURTHER READING

TIMELINE: NINETEENTH- AND TWENTIETH-CENTURY ORIGINS OF COGNITIVE NEUROSCIENCE

	Event	Dates	Contemporaneous developments in the psychological study of cognition*
Chapter 1	Gall promotes his treatise on phrenology	1790–1820s	Philosophy dominates scholarship in psychology
Chapter 1	Flourens performs experiments refuting phrenology-based localization	1820s–1850s	Pioneering experimentation in perception by Weber and Helmholtz gives rise to psychophysics
	Jackson proposes somatotopic organization of motor cortex	1860s	
Chapter 1	Broca, based on work with Tan and other patients, localizes speech to left posterior inferior frontal gyrus.	1863	
Chapter 1	Fritsch and Hitzig demonstrate, with electrical stimulation, the localization of motor functions to posterior frontal cortex	1870	
Chapter 18	Wernicke describes receptive aphasia, thereby localizing speech comprehension	1874	
Chapter 1	Studies of Munk and of Schäfer establish conclusively the localization of visual perception to the occipital cortex	1880s	Wundt founds Institute of Experimental Psychology
Chapter 3	Brodmann publishes cytoarchitectonic map of anatomical organization of the human brain	1909	Dawn of behaviorism
Chapter 3	Berger records human alpha rhythm with extracranial electroencephalography (EEG)	1929	Psychology dominated by behaviorism
Chapter 4	Penfield's electrical stimulation studies map somatosensory homunculus to postcentral gyrus	1930s	

*Note that, although these are not considered in any detail in this textbook, they are provided for historical context.



A BRIEF (AND SELECTIVE) HISTORY

Although the term “cognitive neuroscience” as a moniker for a scientific discipline has only been with us for a few decades, the field has roots that extend back thousands of years. Ancient Egyptians, Greeks, and Romans all had ideas about the corporeal bases of human thoughts and emotions, although many of these did not specify a role for the brain. In preparing the bodies of deceased nobles for the afterlife, for example, ancient Egyptians removed and discarded the brain as an early step in the mummification process. The internal organs that were deemed to be important were preserved in urns that were entombed along with the body. In most ancient civilizations for which there are records, up through and including Roman civilization, the heart was believed to be the organ of thought. By the time we get to Enlightenment-era Europe, however, the central importance of “neuro” for cognition was widely accepted. One highly influential (and, more recently, ridiculed) example was that of German anatomists Franz Josef Gall (1758–1828) and Johann Caspar Spurzheim (1776–1832), who developed a highly detailed scheme, known as **phrenology**, for how the shape of different parts of the skull related to one’s personality and mental capacities. The underlying premise was that the relative bigness or smallness of various parts of the brain would produce convexities or concavities in the overlying skull. A skilled phrenologist, then, could learn something about an individual by palpating that person’s skull. A bulge in the cheekbone below the eye would mean a predilection toward language, whereas an indentation near the left ear would correspond to a relative absence of the trait of “destructiveness” (*Figure 1.1*). One can see how such a scheme, if it had any validity, would have obvious utility for diagnosing maladies of the brain, as well as for assessing personality and aptitude. (Indeed, for a period during the 1800s it was (mis)used in this way quite extensively, particularly in England and in the United States.)

For at least the past 100 years, the psychology and neuroscience communities have viewed virtually all tenets of the phrenological enterprise as being

Tip 1.1

Cognitive neuroscience is grounded in the governing assumption that all cognitive functions arise from the physical, chemical, and physiological properties of the brain and **central nervous system**. Thus, all differences between individuals must reduce to physical factors. In **typically developed** humans, however, these will be microscopic (e.g., connections between **neurons** (*Chapter 2* and *Chapter 10*); varying levels of concentration of chemical transmitters (*Chapter 2*)), and underlain by differences in genetics and/or life experience.

scientifically invalid. From one perspective, we have come to know that subtle, idiosyncratic variations in gross shape from one brain to another have little, if anything, to do with the “kind of person” that one is (however, see *Tip 1.1* for qualification of this point). We also recognize that the assignments of function that Gall gave to various parts of the brain were not based on rigorous science, and turned out to be altogether wrong. A third point, that merits additional elaboration in *Thought Box 1.1*, is that the very selection and definition of functions that phrenologists mapped onto the brain lacked systematicity and rigor. There was, however, at the core of the phrenological enterprise, a powerful idea that has continued to animate many debates about brain function up to the present time – the idea of **localization of function**.

Localization of function vs. mass action

The principle of localization of function refers to the idea that different aspects of brain function, such as visual perception vs. the control of our emotions vs. our talents as musicians, are governed by, and therefore localizable to, different “centers” in the brain. An analogy might be that different functions of the body – extracting oxygen from blood vs. pumping blood vs. filtering blood – are each accomplished by a different organ (i.e., the lungs, the heart, and the kidneys) that

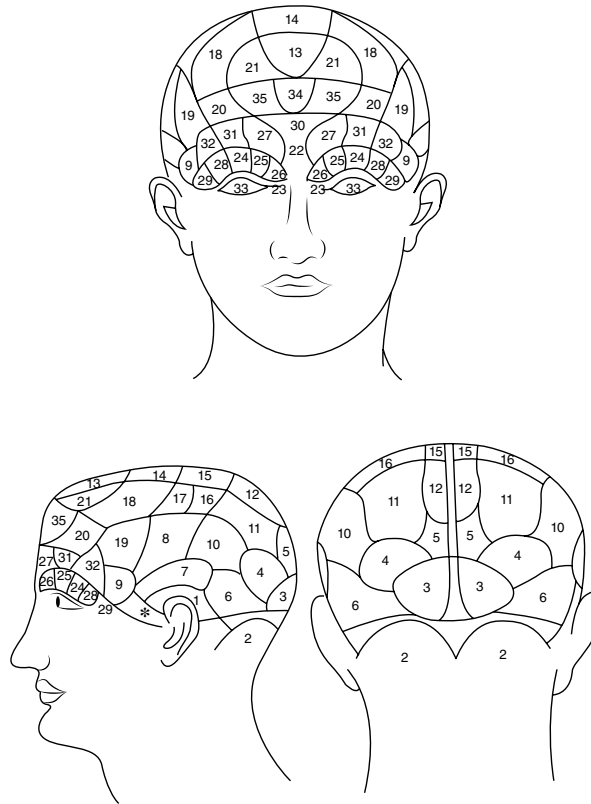


FIGURE 1.1 A phrenological map of the anatomical organization of mental faculties. They are organized into two primary categories (**bold**), with subcategories (underlined) under each. The first primary category is **Affective Faculties**: Propensities: * – Alimentiveness; 1 – Destructiveness; 2 – Amativeness; 3 – Philoprogenitiveness; 4 – Adhesiveness; 5 – Inhabitiveness; 6 – Combaticiveness; 7 – Secretiveness; 8 – Acquisitiveness; 9 – Constructiveness; Sentiments: 10 – Cautiousness; 11 – Approbativeness; 12 – Self-Esteem; 13 – Benevolence; 14 – Reverence; 15 – Firmness; 16 – Conscientiousness; 17 – Hope; 18 – Marvelousness; 19 – Ideality; 20 – Mirthfulness; 21 – Imitation. The second primary category is **Intellectual Faculties**: Perceptive: 22 – Individuality; 23 – Configuration; 24 – Size; 25 – Weight and Resistance; 26 – Coloring; 27 – Locality; 28 – Order; 29 – Calculation; 30 – Eventuality; 31 – Time; 32 – Tune; 33 – Language; Reflective: 34 – Comparison; 35 – Causality. Source: Spurzheim, 1834.

are located in different parts of the body. This notion can be contrasted with an alternative idea, **mass action**, according to which a particular function can't necessarily be localized to a specific area of the brain and, conversely, any given area of the brain can't be thought of as a "center" that is specialized for any one function. To stick with our analogy to familiar parts of the body below the neck, we can illustrate the principle of mass action by zeroing in on the kidney. The overall function of the kidney – filtering blood – is

carried out in the same way by the top portion and the middle portion and the bottom portion. To understand how the kidney does its job, one could study in detail the inner workings of only the top, or only the middle, or only the bottom of this organ, and one would learn the same thing from each. In effect, then, different zones of the kidney are "interchangeable" with respect to understanding their function. Now project yourself back in time a few centuries to a time when what I've just written hasn't yet been

THOUGHT BOX

1.1 What's a function? (And what qualities make the neural bases of a function amenable to experimental study?)

A quick inspection of an iconic phrenological bust, such as that illustrated in *Figure 1.1*, often elicits giggles: *Acquisitiveness*? *Conscientiousness*? *Amativeness*? On what basis can one assert that any of these labels actually correspond to a unitary, discrete “function”?

Near the end of this chapter we'll note that the functions of motor control and vision are relatively easy to observe and to measure. The fact that they also turn out to be localizable with a high degree of anatomical specificity reinforces the idea that each can be considered a discrete function of the brain. But can the same be said about, say, *conscientiousness*? It is true, of course, that an organism without a brain cannot exhibit conscientiousness, and, therefore, the phenomenon could not exist without a brain. However, might it not be the case that “conscientiousness” is just a label that we, as denizens of highly organized societies, have given to a certain collection of attributes that doesn't correspond to any single discrete mental capacity? For example, if a student sends me a thank-you note for writing a letter of recommendation, I will consider her to have displayed conscientiousness. But could it not be the case that she was conditioned during her upbringing to seek positive reinforcement from her parents (*You're such a good girl for writing those thank-you notes!*), and that her note to me is “merely” the product of an association that she has formed between writing a thank-you note and this reinforcement? Were this the case, it wouldn't be possible to localize conscientiousness, per se, and any attempt to do so would be destined to yield erroneous conclusions.

This exercise illustrates a fundamental shortcoming of phrenology: many, if not all, of the “functions” that it sought to map onto the brain were simply not valid, because they were derived from nothing more than Gall's intuition. More generally, it illustrates two principles that are highly relevant to contemporary cognitive neuroscience. The first is that a model of the neural instantiation of a cognitive function depends critically on the validity of the function that it seeks to explain. The question of **construct validity** will be important for every domain of behavior that we consider in this book. For many, the formal models of the construct under study will come from one of the “different scholarly disciplines” invoked in the Introduction to Section I. A second principle is that not all aspects of cognition and behavior are equally amenable, with current theories and methods, to a neural level of explanation. To illustrate, we can return again to conscientiousness. It so happens that in the field of personality psychology, *conscientiousness* is a valid construct, one of the “big five” dimensions along which an individual's personality can vary (along with *agreeableness*, *extroversion*, *openness*, and *neuroticism*). Each of these is formalized as a statistical factor to which many traits contribute. (For *conscientiousness*, these include the extent to which one “follows a schedule,” “forgets to put things back in their proper place,” etc.) However, none of this alters the reasoning from the preceding paragraph, which is that not all describable traits are easily reduced to a neural level of analysis.

discovered. It is an era when biomedical research techniques are limited, and the best tool that you have for studying the function of an organ is to damage a portion of it and then observe the consequent impact of this damage on its function. Your kidney research

would indicate that damaging comparable-sized regions of upper vs. middle vs. lower kidney has the same effect in all cases: an overall decline in the efficacy of blood filtration. Thus, you would have discovered that a principle of mass action applies to the

Tip 1.2

Some scientists reserve the use of *neuropsychology* to refer to studies with humans who have sustained damage to the brain, and refer to analogous experiments with nonhuman animals simply as *lesion studies*.

kidney: a larger **lesion** results in a larger diminishment of the rate of blood filtration; smaller lesions result in a smaller diminishment of the rate of blood filtration; and, critically, the location of the damage doesn't seem to matter.

Now, let's return to the functions of the brain. In the decades following the introduction of phrenology, and in some instances in reaction to it, scientists and physicians began pursuing the idea of localization of function in the brain with methods reflecting the maturation of the scientific method that was occurring in many branches of science, from biology to chemistry to physics. At this general level, this entailed the *a priori* articulation of falsifiable hypotheses (i.e., stating the hypothesis prior to performing the experiment), and the design of controlled laboratory experiments that could be replicated in other laboratories. An important advance for studies of the brain, in particular, was the careful analysis of the behavioral consequences resulting from damage to a particular brain structure. This method has come to be known as **neuropsychology** (*Tip 1.2*). Armed with this approach, specific localizationist claims of phrenologists began to be disproven. Perhaps most influential were the studies of French scientist Pierre Flourens (1794–1867). A tireless critic of phrenology, Flourens did much of his experimental work with pigeons and with dogs. This research was influential on two levels. First, particular experiments of Flourens disproved specific phrenological claims, such as his demonstration that damage to the **cerebellum** disrupted locomotor coordination, but had no effect on amateness (i.e., sexual arousal), as Gall's model would have predicted (see *Figure 1.2*; and *Tip 1.3*). Secondly, and at a broader level, Flourens' studies of the cerebral cortex largely failed to find evidence for localization of function. Thus, although damage to

Tip 1.3

Gall's formal study of localization of function was not only controversial in scientific circles, but also had political implications. As the historian of science Stanley Finger puts it, "Authorities of the Austrian Catholic church viewed this work as championing materialism, atheism, and fatalism bordering on heresy." As a result, the Holy Roman Emperor (who ruled Vienna, where Gall had launched his career) sent to Gall a threatening letter that read, in part, "This doctrine concerning the head ... will perhaps cause a few to lose their heads" (Finger, 2000, pp. 124–125). Within a few years, Gall (and his head) left Vienna permanently, eventually settling in Paris.

the cortex invariably produced marked disruption of behaviors associated with judging, remembering, and perceiving, these impairments seemed to occur regardless of what part of the brain had been damaged. By inference, such results seemed to indicate that all regions of the cortex contributed equally to these behaviors. (Note that the same was not true of Flourens' studies of the brainstem, to which, for example, the cerebellum belongs (see *Figure 1.2* and *Web Link 1.1*).)

Another important concept to come out of the work of Flourens was derived from the fact that, over time, animals with experimental damage to a part of cortex often recovered to presurgical levels of functioning. Because this occurred without evident repair of the damaged tissue itself, it was assumed that intact areas of the brain had taken over this function. This gave rise to the concept of **equipotentiality**, the idea that any given piece of cortical tissue had the potential to support any brain function.

Roughly 50 years prior to the writing of this textbook, and 130 years after the heyday of phrenology, neuroscientists Charles Gross and Lawrence Weiskrantz wrote that "The 'heroic age of our field' was opened by Gall (1835) ... [who] stimulated the search for centers and gave the mass action-localization pendulum its first major swing" (Gross and Weiskrantz, 1964). Implied in this

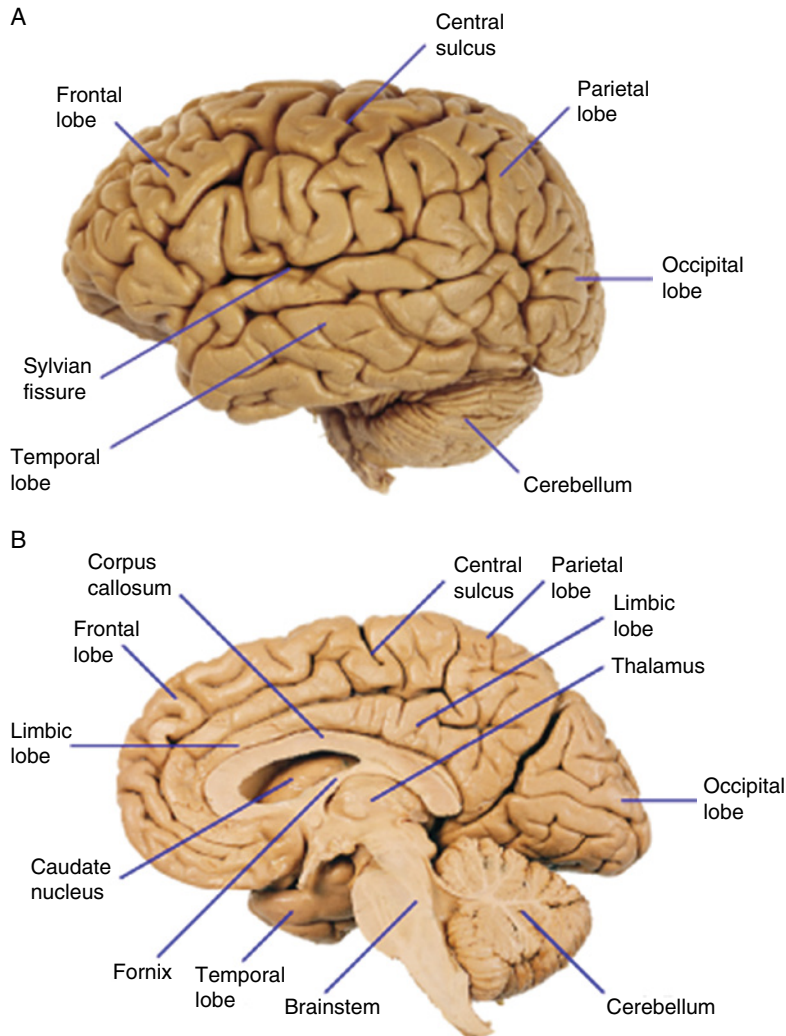


FIGURE 1.2 The human brain (see *Chapter 2* and *Web Link 1.1*, a 3-D rotatable and “dissectable” human brain, for labels and terminology). **A.** Lateral view. **B.** Medial view. Source: Dr Colin Chumbley / Science Photo Library. Reproduced with permission.

quote was that understanding the localization–mass action dichotomy would provide insight into understanding key contemporary problems in neuroscience. Indeed, *Thought Box 1.2* elaborates further on this theme precisely because it retains its relevance today, and will prove to be useful for understanding many of the concepts and controversies that are prominent in contemporary cognitive neuroscience.

The first scientifically rigorous demonstrations of localization of function

Although the concepts advocated by Gall, on the one hand, and Flourens, on the other, still resonate today, the same can not be said for most of the “facts” that their work produced. Rather, it was in the mid to late 1800s, during what can be seen as the first return of

THOUGHT BOX

1.2 Localization vs. mass action down through the years

The core question in cognitive neuroscience is *How does the functioning of the brain give rise to cognition and behavior?* One way to approach the question of how a system works is to divide its overall behavior into logically discrete functions, and to determine whether different physical parts of the system might differentially support the different functions. To determine how an automobile works, for example, one might start by logically “decomposing” its overall behavior into two plausibly separable functions, accelerating and sustaining movement vs. decelerating and stopping. This approach naturally leads to a way of thinking about the system as being made up of discrete parts, each of which carries out a different function. However, not all systems are made up of discrete parts, each with a separate function, in the way that an automobile is. One example of such a system is a school of fish. It has properties (overall shape, smoothness of outside surfaces, changes of speed and direction) that cannot be localized to component parts. Each of these properties can be said to be an **emergent property**, in that they don’t exist in any of the constituent elements (i.e., in any one fish), but they “emerge” at the level of the school. (We shall see that many cognitive functions, including consciousness, are believed by many to be emergent.) When considering the school of fish, all parts perform the same function and are interchangeable (i.e., they display equipotentiality), and removing 20% of the fish from one region of the school vs. from a different region will have identical effects on the school (i.e., the principle of mass action applies). Localization of function is not possible with a school of fish. (Note, however, that localization of functions within an individual fish (e.g., moving the tail vs. extracting oxygen from water) *is* possible, and so specifying the grain of detail at which one is studying a system is very important.)

Because automobiles are manmade, we don’t actually have to perform experiments to determine how they work. And it is precisely because we know how they are designed and built that we know that a localization-of-function perspective is the correct way to think about how the different functions of the car are accomplished. Because a school of fish is fully observable (i.e., we know effectively everything that there is to know about what it’s made of), we know that the properties of emergence, equipotentiality, and mass action apply to it. (These are often summarized under the rubric **distributed systems**.) The brain is different from these two examples, however, in that we don’t know all there is to know about how it is “made.” To study it, therefore, cognitive neuroscientists devise models, then test these models with experiments. Although it is not always explicitly acknowledged, these models and/or the techniques used to test them often make an assumption that the brain system under investigation works on fundamentally localizationist or distributed principles. As we start to consider different contemporary problems in cognitive neuroscience, keep in mind that the choice of model and/or method can bias the extent to which one is likely to draw localizationist or mass-action/emergence conclusions from one’s experimental data.

the pendulum back toward localization, that we see the emergence of principles of brain function that, at least to a first order of approximation, have held up through to the present day. These involved the functions of motor control, vision, and language.

The localization of motor functions

Beginning in the 1860s, British neurologist John Hughlings Jackson (1835–1911) described the systematic trajectory of certain focal **seizures** that start in the fingers and spread along the arm toward the

trunk, sometimes ending with a loss of consciousness. From this distinctive pattern, which has since come to be known as the *Jacksonian march* (it's as though the seizure is marching along the body), Jackson proposed that the abnormal brain activity presumed to be the cause of the seizures begins in a part of the cortex that controls the fingers, then moves continuously along the surface of the cortex, progressively affecting brain areas that control the palm of the hand, the wrist, the forearm, and so forth. There were two important implications of Jackson's theory. The first was quite simply the proposal that the capacity for movement of the body (i.e., motor control) is a function that is localized within the brain. The second was what has come to be understood as a fundamental principle underlying the organization of function within many portions of the brain, which is that the organization of the brain can mirror the organization of the body (or, as we shall see, of a particular part of the body). Specifically in this case, the proposal was that the area of the brain that controls the muscles of the fingers is adjacent to the area of the brain that controls the muscles of the palm, which is adjacent to the area of the brain that controls the muscles of the wrist, and so forth. Thus, the functions of what came to be known as the *motor cortex* are physically laid out on the surface of the brain in a kind of map of the body (that is, in a **somatotopy**). In this way, the idea of a lawful, **topographic organization** of function was introduced in the void left by the (by-now-largely-discredited) arbitrary, willy-nilly scheme of the phrenologists. (The principle and characteristics of somatotopy will be considered in detail in *Chapter 4* and *Chapter 7*.)

Although the ideas that Jackson proposed were based on careful observation of patients, an idea such as the somatotopic organization of motor cortex couldn't be definitively evaluated without either direct observation or manipulation of the brain itself. The ability to undertake such definitive empirical investigation became possible because Age-of-Enlightenment advances in thinking about how science should be conducted, such as the importance of the scientific method, were being paralleled by technical advances that afforded improved experimental methods. Of particular importance was the development of methods for performing aseptic (i.e., sterile) surgery. These enabled experimenters to keep animals alive for weeks

or longer after performing the **craniotomy** that was necessary to create a lesion or to manipulate the functioning of the brain. Prior to this, infection would often limit postsurgical survival to a matter of hours or days. This technical advance had several important consequences, one of which was that it opened the way for direct stimulation of, and subsequently recording from, the brain of an intact animal (techniques that fall under the category of **neurophysiology**; see *Methodology Box 1.1*). Thus it was that German physicians Gustav Fritsch (1838–1927) and Eduard Hitzig (1838–1907) reported that electrical stimulation of **anterior** portions of the cerebral cortex of the dog, in the **frontal lobe**, produced movements on the opposite side of the body (Fritsch and Hitzig, 1870). Noteworthy were two facts. First, comparable stimulation of a more **posterior** brain region, the parietal lobe, did not produce body movements (*Figure 1.3*). Based on this demonstration of the **anatomical specificity** of their effect (see *Methodology Box 1.1*), Fritsch and Hitzig explicitly challenged the idea of equipotentiality that had been advocated by Flourens. Second, the part of the body affected by electrical stimulation (e.g., neck, forelimb, hindlimb) varied systematically with the positioning of the stimulating electrode. This very directly supported Jackson's idea of a somatotopic organization of motor functions in the brain.

The localization of visual perception

A second function that was the focus of intense scientific investigation during this period was visual perception. Here, too, the work of Flourens had been influential. Although it had brought to light the principle of crossed lateralization of function – in that lesions of the left hemisphere produced visual impairments in the right visual field, and vice versa – it had found no evidence that this general pattern varied as a function of *where* in the hemisphere the lesion was made. That is, Flourens had failed to find evidence for localization of visual function within a **cerebral hemisphere**. As with other of his null findings, however, this one was overturned thanks to the advent of newer, better, empirical methods. In the case of vision research, the refinement of methods of experimental lesioning was the critical development that led to an important discovery. In particular, the development of aseptic surgical techniques resulted in longer and

METHODOLOGY BOX

1.1 Neuropsychology and the importance of anatomical specificity

Neuropsychology relies on the logic that one way to understand the workings of a system is to systematically remove (or inactivate) parts of it and observe how the removal of each part affects the system's functioning. A neuropsychological study can address the question *Does region A make a necessary contribution to behavior X?*, but it cannot directly address how it is that *region A* may be doing so. That is, it cannot address questions of mechanism. Neurophysiological approaches, on the other hand, can study the manner in which the function of *region A* gives rise to *behavior X*. To give a concrete example, *Chapter 1* describes how neuropsychological experimentation was used to determine definitively that visual perception is localized to the occipital cortex. Not until *Chapter 3*, however (and, chronologically, not until roughly 80 years later), will we see that electrophysiological experiments revealed how it is that neurons in the occipital lobe process visual information. At another level, however, a carefully conducted neuropsychological experiment can support stronger inference about a region's contribution to behavior than can experiments that measure neurophysiological variables, because the former can tell us definitively about whether or not a region's contribution to a type of behavior is *necessary*. Neurophysiological measurement studies, in contrast, are inferentially limited to demonstrating correlations between activity in the brain and the behavior that is being studied. Neurophysiological stimulation studies, such as those of Fritsch and Hitzig that are considered in this chapter's section on *The localization of motor functions*, fall somewhere in between.

In both types of experiment, the strength of one's conclusions can depend on anatomical specificity. What made the Fritsch and Hitzig (1870) experiment so powerful, for example, was the specificity with which it demonstrated that stimulation of a particular portion of the frontal lobe in the left hemisphere produced movement of the right forelimb, and not the right hindlimb, nor the left forelimb. Further, stimulating a nearby area selectively produced movement of a different part of the body, the right hindlimb. Had the authors limited their report to the initial demonstration, two important questions would have been left unanswered. First, *might stimulation of other regions also produce movement of the right forelimb?* Hinging on the answer to this would be an understanding of how localized is the control of the right forelimb. Second, *does stimulation of this region of the brain produce movement of other parts of the body?* The answer to this would give an indication of how specific is the function of brain area in question. These considerations can be particularly important in neuropsychological studies of humans, such as those described in this chapter's section on *The localization of speech*, because such studies often rely on "accidents of nature," such as stroke, **neurodegenerative disease**, or head injury. In such studies, scientists have no control (other than patient selection) over the extent of damage in such patients, neither in terms of how many structures may be damaged, nor of the overall volume of tissue that is affected. When patient lesions are large, there is an inherent difficulty in determining which of the damaged structures is responsible for the observed behavioral deficit. One logical way to tackle this problem, the "double dissociation of function," will be introduced in *Chapter 5* (in *Research Spotlight 5.1*. and *Methodology Box 5.1*).

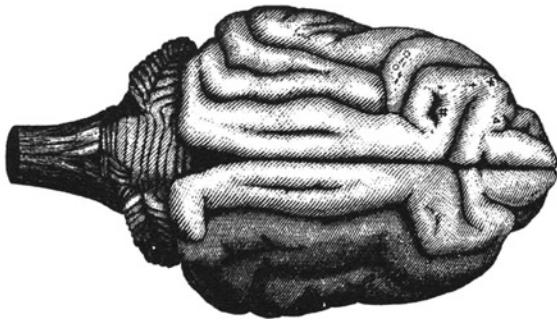


FIGURE 1.3 Illustration of a top-down view of a dog's brain, with symbols indicating, from rostral to caudal, the areas whose stimulation produced muscle contractions on the right side of the body of the neck, forelimb, and hindlimb, respectively. Source: Fritsch and Hitzig, 1870.

healthier postsurgical survival of experimental animals than had previously been possible. This, in turn, afforded considerably more sophisticated and conclusive assessments of behavior. Indeed, for a time, it was not uncommon for a researcher to bring a lesioned animal to a scientific conference so as to demonstrate the behavioral alterations produced by a particular lesion. Perhaps for related reasons, this period also saw an increase in research with monkeys, whose brain and behavior are more comparable to those of humans than are those of, say, birds and carnivores. It is much more expensive to acquire and house monkeys relative to other species, such as birds, rodents, and carnivores. Thus, research with monkeys would not have been considered practical prior to nineteenth-century refinements of surgical techniques. (See *Thought Box 1.3* for a consideration of the role of nonhuman animals in neuroscience research.)

The discovery of a region specialized for motor control was quickly followed by intense research on the neural bases of visual perception. Using electrical stimulation techniques and surgical lesioning, initial research in dogs and in monkeys indicated a privileged role for posterior regions in supporting visual perception. That is, whereas it had been shown conclusively that motor control did *not* depend on posterior regions, the opposite was true for vision. The final decades of the nineteenth century witnessed vociferous debate about whether the “visual centres,” a characterization of British physiologist David Ferrier (1843–1928), were localized to a region of

the parietal cortex or to a region of the occipital cortex (*Figure 1.4*). This research, carried out primarily in monkeys and most prominently by Ferrier, by Berlin-based physiologist Hermann Munk (1839–1912), and by London-based physiologist Edward Schäfer (1850–1935), gradually converged on the conclusion that the region whose destruction produced frank and lasting blindness – as opposed to more nuanced and transient visual deficits – was the occipital cortex. This conclusion, reinforced by observations of human patients with head injuries, led to the universally accepted localization of **primary visual cortex** to the occipital lobe.

The localization of speech

The last function of the brain that we will consider in this introductory chapter is language; more specifically, the ability to speak. The faculty of language had been localized by phrenologists to a region of the frontal lobe directly behind the eyes. (This idea is said to have derived from Gall's observation of a classmate who had a prodigious verbal memory and protruding eyes, the latter presumed by Gall to be the result of a bulging frontal lobe that pushed the eyes forward.) As was the case with motor control, the post-phrenology study of language began with clinical observations. In the decades prior to the 1860s, isolated reports indicated, in some cases, that damage to the left side of the brain was associated with impairments of speech, in others, that damage to anterior portions of the brain had this effect. As these cases accumulated, the more specific idea emerged that anterior portions of the left hemisphere were important for speech. During the 1860s, French surgeon Paul Broca (1824–1880) published a series of case studies that confirmed this idea. Most celebrated was the case of the **stroke** patient “Tan,” so nicknamed because this was the only sound that he could make with his mouth. (Phonetically, “tan” is pronounced in French as /tân/.) Importantly, Tan's impairment was specific to *speech production* (i.e., talking), because he could understand and follow verbal instructions that he was given. Additionally, there wasn't any obvious paralysis of the speech apparatus, in that, despite some right-side-of-the-body motor impairment, he could eat and drink and make his famous verbal utterance. Upon Tan's death, Broca examined the patient's brain and concluded from the prominent damage that he observed (*Figure 1.5*) that

THOUGHT BOX

1.3 The role of nonhuman animals in neuroscience research

It is evident from this historical overview that modern cognitive neuroscience could not have been developed without research conducted on nonhuman animals. It is equally true that research performed on nonhuman animals continues to play a critical and necessary role in every branch of contemporary cognitive neuroscience. The reason for this, as you will see in each chapter, is that many questions about physiology and anatomy cannot be answered scientifically without carrying out experiments that are invasive. Reading this book will also make clear that, although we have learned an astounding amount of information about how the workings of the brain give rise to cognition, there remains an enormous amount that we still don't know. Despite recent and ongoing advances in a variety of noninvasive methods that can be used to study the brain (such as neuroimaging and computer modeling), there will continue, for the foreseeable future, to be questions that can be answered through no means other than invasive experimentation.

Because of these facts, another sociological factor with which the cognitive neuroscientist must be conversant (in addition to those raised in the second paragraph of this chapter) is the ethics of animal research. This is true even for researchers such as the author of this textbook, whose research is conducted exclusively on human subjects. This is because many ideas that motivate our experiments come from research that was performed with animals, and our ability to interpret our data often depends on what has been learned from animal research. For an informal index of this, consider the syllabus for my research group's lab meetings. During the academic semester that preceded the writing of this thought box, my group met on 14 occasions, and on each we discussed an article from a scientific journal that is relevant for our research. Of these 14 meetings, four (i.e., nearly 30%) were devoted to articles describing research performed with nonhuman animals.

This brings up a stylistic point: throughout the book, when I refer to "primates" without any other context, I'm referring to all species within this taxonomic order (which includes humans). Typically this will be in the context of a structure/function/behavior that is shared by primates, but does not generalize to other mammals.

The ethics of animal research are complex, and the topic can arouse strong emotions. Debates are often framed in hypothetical terms for which there are no objectively "right" or "wrong" answers. One question that has been contemplated, for example, is *If one encounters a burning building and there's only time to save the human who is inside or the mouse that is inside (or the dog, or the monkey, ...), which does one save?* You can see how two people could devote an afternoon to debating whether or not this analogy is even appropriately relevant to the question of animal research. Although a more detailed consideration of this topic is outside the purview of this book, the *Further Reading* from neuroscientist Dario Ringach (2011) nicely summarizes many of the arguments in favor of and in opposition to animal research, and provides references to writings that represent both the pro and the con sides of the debate.

the ability to speak was localized to the "posterior third of the third convolution" (a portion of the left inferior frontal gyrus that today is known as Broca's area). It's worth taking a moment to consider why Broca, rather than some of his predecessors and contemporaries who had made similar observations

of patients, gets the credit for this discovery. One important factor is that Broca's reports were seen as confirming an idea that had been predicted ahead of time – a sequence of events that fit with the era's increasing emphasis on hypothesis testing as an important element of the scientific method.

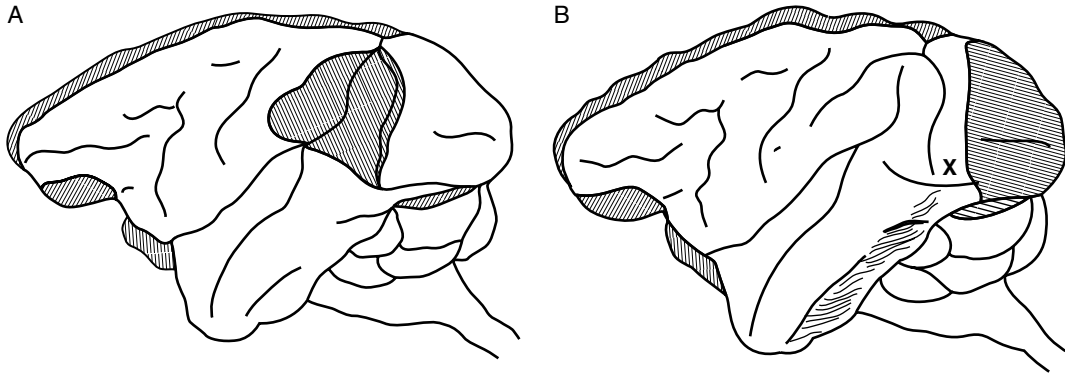


FIGURE 1.4 Two experimental lesions produced by Ferrier in experiments designed to localize the cortical locus of visual processing. Panel **A** illustrates a lesion of the parietal lobe that Ferrier interpreted in 1876 as producing lasting blindness, and, a decade later, as producing only temporary blindness. Panel **B** illustrates a lesion of the occipital lobe that he interpreted as producing no visual impairments. Neuroscientist Charles Gross (1998) has marked with an X the region that we now know to represent central (i.e., foveal) vision, a topic to be explored in depth in *Chapter 3*. These lesions, and their interpretation, illustrate how difficult it was for nineteenth-century neuroscientists to localize cognitive functions in the brain.

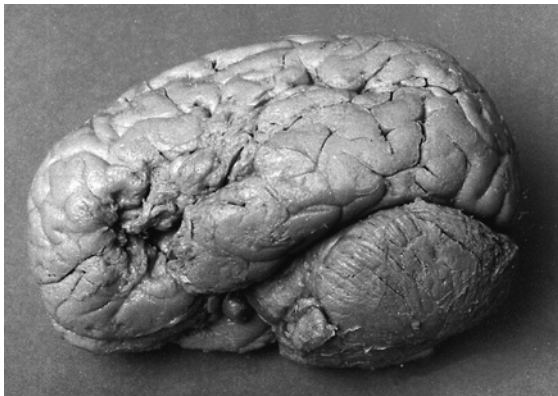


FIGURE 1.5 The left external surface of the brain of Broca's patient Tan. Note the extensive damage in the posterior inferior frontal gyrus, a region that has come to be known as Broca's area. Source: Reproduced with permission from the Dupuytren Museum, Pierre and Marie Curie University, Paris. Reproduced in N. F. Dronkers, O. Plaisant, M. T. Iba-Zizen, and E. A. Cabanis. Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong *Brain* (2007) 130 (5): 1432–1441, Figure 3A. Photograph by Bruno Delamain.

WHAT IS A BRAIN AND WHAT DOES IT DO?

As we conclude this whirlwind review of the birth of modern brain science, it is interesting to consider why *motor control*, *vision*, and *language* were among the first three functions of the brain to yield to newly developed ways of thinking about and doing science. With particular relevance to the first two, let's pose the general questions of *What is a brain?* and *What does it do?* For example, what properties do animals with brains share that living organisms without brains (trees, for example) do not? One answer is that brains confer the ability to detect changes in the environment – let's say a falling rock – and to take an appropriate action: to get out of the way. The hapless tree, on the other hand, has no means of acquiring the information that something potentially dangerous is happening, nor the ability to do anything about it. So it gets crushed. Thus, the ability to see (i.e., vision) and the ability to move (i.e., motor control) are functions that are relatively easy to observe and measure. The same is not true for more nuanced examples of brain-mediated events that

occur in the world. Let's take the example of writing the textbook that you're reading right now. Certainly, vision and motor control were involved. (For example, when I received an emailed inquiry from the publisher about writing this textbook, I used vision to read the pattern of black and white characters on the computer screen that conveyed the inquiry from the publisher.) There were also, however, many additional steps. These included judgment and decision making (*Is it worth all the extra work to take this on? Are the terms of my contract with the publisher acceptable?*), retrieving long-term memories (*There's that great quote from Gross's chapter in the Warren & Akert book that I should use here*), making (and retrieving) new long-term memories (*Where did I put my copy of the MIT Encyclopedia of the Cognitive Sciences the last time I was working on this chapter?*), and much, much, more. These latter operations are "internal" or "mental," in the sense that they happened in my mind, without any obvious way of seeing or measuring them. Thus, although many aspects of so-called **high-level cognition** – including decision making and different aspects of memory – are the focus of intense research in contemporary cognitive neuroscience, we shall see that the ability to define what constitutes a distinct function (analogous to vision or motor control), as well as the ability to localize the brain region(s) on which these functions depend, become much more complicated propositions. The principles that explain how such functions work will be less accessible to direct observation and measurement than are brain functions that are more closely tied to either the "input" of information to the brain (in the example of vision), or the "output" of the brain that is expressed via movements of the body (i.e., the actions of the organism). (However, as we shall see, there are influential perspectives on brain function as a whole that construe it as a hierarchically organized series of **sensorimotor circuits**, with the more abstract ones (e.g., all the processes that went into writing a textbook) having been superimposed, over the course of evolution, onto more basic ones (e.g., reflexive avoidance of a threat).)

Okay, so the neural bases of vision and motor control may have been among the first to be studied "scientifically" because they are easy to observe. But the

same cannot be said about language. Although its production clearly has motoric components, we've already reviewed ways in which the language impairment experienced by Tan and other patients was fundamentally different than paralysis. Indeed, in ways, language can be construed as being at the opposite end of the concrete-to-abstract continuum of human faculties that spans from such "concrete" (that is, easily observable and relatively easily measurable) functions as vision and motor control to the most abstract aspects of conscious thought. Although language is readily observable, there is also a sense in which it epitomizes abstract, high-level cognition. One reason is that language entails the use of arbitrary, abstract codes (i.e., a **natural language**) to represent meaning. Intuitively, because many of us have the sense that we "think in words," language can be viewed as "the stuff of thought," and thus epitomizes an "internal," indeed, a *cognitive*, function. And so I'm reticent to conclude that language was among the first human faculties to be studied with modern neuroscientific techniques because it is easy to observe. Instead, what I'll offer is that it is language's intuitive "specialness," and its seeming uniqueness to humans, that has made it a focus of interest for millennia. The Age of Enlightenment, then, may simply be the epoch of human history when developments in the scientific method caught up with an age-old focus of human curiosity.

LOOKING AHEAD TO THE DEVELOPMENT OF COGNITIVE NEUROSCIENCE

To conclude this introductory chapter, we see that by the dawn of the twentieth century, brain scientists were beginning to investigate the brain from both ends of the continuum that captures its functions in relation to the outside world. In the process, we can think of modern brain science as developing along two paths. The first, pursued by physiologists and anatomists who studied brain structure and function in nonhuman animals, focused on neural *systems* that support various functions (e.g., visual perception and motor control), and has come to be known as systems neuroscience. The second took as its starting point a human behavior, such as language, and proceeded with the logic that careful

study of the way that the behavior in question is altered by insults to the brain can be informative, with respect both to how the behavior is organized, and to how the workings of the brain give rise to this organization. This latter approach has matured into the allied disciplines of behavioral neurology (when carried out by physicians) and neuropsychology (when carried out by non-physician scientists). Now of course, the notion of two categories of brain science developing in parallel at the dawn of modern brain science is an overly facile dichotomization imposed by this author, and many instances can be found that do not fit neatly into such a taxonomy (see *Tip 1.4*). Nonetheless, we will find this distinction to be useful as we proceed. It is these two scientific traditions summarized in this chapter, together with a third, experimental psychology (which also got its start in the second half of the nineteenth century, but which won't be covered in depth here (see *Further Reading* sections)), that provide the foundation from which modern cognitive neuroscience has emerged. Before we plunge full-on

Tip 1.4

One example of the two major categories of neuroscience influencing the work of one investigator can be seen with Eduard Hitzig. His electrical stimulation experiments with Frisch are recognized as foundational in launching systems neuroscience. A formative influence that may have inspired these experiments, however, was the time earlier in his life when Hitzig served as a battlefield physician in the Prussian army. In this capacity, he encountered many soldiers who evinced various behavioral deficits as a result of injuries to the head acquired in battle.

into cognitive neuroscience itself, however, we need to review some facts about the brain – how it is put together (gross anatomy), and how it works (cellular physiology and network dynamics). These will be the focus of *Chapter 2*.

END-OF-CHAPTER QUESTIONS

1. How does cognitive neuroscience differ from the related disciplines of cognitive psychology and systems neuroscience?
2. Although most of the specific claims arising from phrenology turned out to be factually incorrect, in what ways did they represent an important development in our thinking about the brain?
3. One of the major flaws of phrenology concerns the functions that it sought to relate to the brain: in some cases, they were not valid constructs; in others, they weren't amenable to neuroscientific explanation at the level that Gall was seeking to capture. From the phrenological bust in *Figure 1.1*, select at least one function to which each of these critiques applies, and explain your reasoning.
4. In the nineteenth century, what kind of scientific evidence was marshaled to support mass-action models of the brain? To support localizationist models of the brain?
5. Might it be possible that some tenets of mass-action and localization-of-function models could both be true? If yes, how might this be possible?
6. How does the concept of anatomical specificity relate to testing hypotheses of mass action and localization of function?
7. The neural bases of what domains of behavior were the first to be systematically explored in the second half of the nineteenth century? For each, what is a likely explanation?
8. What is the principle of topographic representation in the brain? Although this chapter emphasized the topographic organization of the motor system, according to what principles/dimensions might the topography of various sensory modalities (e.g., vision, somatosensation, audition) be organized?
9. Apart from one study that employed electrical stimulation, this chapter described experiments that relied on inferring the neural bases of a function from observation/measurement of the consequences of damage to different parts of the brain. Nonetheless, there are two fundamentally different ways in which one can pursue such research. Describe them, and name the disciplines of science and/or medicine that are associated with each. Which of the two requires the use of nonhuman experimental animals?

REFERENCES

- Finger, Stanley. 2000. *Minds Behind the Brain*. Oxford: Oxford University Press.
- Fritsch, Gustav, and Eduard Hitzig. 1870. "Über die elektrische Erregbarkeit des Grosshirns." *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin* 37: 300–332. [Available in English translation as: Fritsch, Gustav, and Eduard Hitzig. 2009. "Electric excitability of the cerebrum (Über die elektrische Erregbarkeit des Grosshirns)." *Epilepsy and Behavior* 15 (2): 123–130. doi: 10.1016/j.yebeh.2009.03.001.]
- Gross, Charles G. 1998. *Brain, Vision, Memory*. Cambridge, MA: MIT Press.
- Gross, Charles G., and Lawrence Weiskrantz, L. 1964. "Some Changes in Behavior Produced by Lateral Frontal Lesions in the Macaque." In *The Frontal Granular Cortex and Behavior*, edited by John M. Warren and Konrad Akert, 74–98. New York: McGraw-Hill.
- Spurzheim, Johann C. 1834. *Phrenology or the Doctrine of the Mental Phenomenon*, 3rd ed. Boston: Marsh, Capen, and Lyon.

OTHER SOURCES USED

- Calvin, William H., and George A. Ojemann. 1994. *Conversations with Neil's Brain*. Reading, MA: Addison Wesley.
- Finger, Stanley. 1994. *Origins of Neuroscience*. Oxford: Oxford University Press.
- Gross, Charles G. 2007. "The Discovery of Motor Cortex and Its Background." *Journal of the History of the Neurosciences: Basic and Clinical Perspectives* 16 (3): 320–331. doi: 10.1080/09647040600630160.
- Holyoak, Keith J. 1999. "Psychology." In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil, xxxix–xlix. Cambridge, MA: MIT Press.
- Kolb, Bryan, and Ian Q. Whishaw. 2003. *Fundamentals of Human Neuropsychology*, 5th ed. New York: Worth Publishers.
- Whitaker, Harry, A. 1999. "Broca, Paul." In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil, 97–98. Cambridge, MA: MIT Press.

FURTHER READING

- Boring, Edwin G. 1929. "The Psychology of Controversy." *Psychological Review* 36 (2): 97–121.
A classic overview of key developments in nineteenth-century psychology and psychophysics, adapted from Boring's Presidential Address to the American Psychological Association, delivered on 28 December 1928. Boring was a professor at Harvard University.
- Finger, Stanley. 2000. *Minds Behind the Brain*. Oxford: Oxford University Press.
A history, stretching from ancient Egypt to twentieth-century North America and Europe, of scientists who have made seminal contributions to the understanding of the structure and functions of the brain. It is longer and covers a broader scope of topics than Gross (1998). Finger has authored many authoritative books on the history of neuroscience, and is Editor-in-Chief of the Journal of the History of the Neurosciences.
- Gross, Charles G. 1998. *Brain, Vision, Memory*. Cambridge, MA: MIT Press.
A highly engaging collection of "Tales in the history of neuroscience." It spans the same range of human history as does Finger (2000), but with fewer chapters and a narrower focus of the functions named in the title. The author is, himself, a highly respected visual neuroscientist, now Emeritus Professor at Princeton University.
- Ringach, Dario L. 2011. "The Use of Nonhuman Animals in Biomedical Research." *American Journal of Medical Science* 342 (4): 305–313.
Ringach is a UCLA-based neuroscientist whose person, family, and house have been the targets of harassment and intimidation by animal rights activists. This fact, however, does not color this even-toned consideration of this debate. Among other things, he urges "civil discourse ... free of threats and intimidation," and concludes that "The public deserves an open and honest debate on this important topic." The article contains extensive references to original writings by individuals offering many different perspectives on this debate.

CHAPTER 2

THE BRAIN

KEY THEMES

- The central nervous system (CNS) is organized into four parts: the spinal cord; the brainstem; the cerebellum; and the cerebrum.
- The computational elements of the brain are its neurons, whose cell bodies are primarily located in cortex, on the outside surfaces of the cerebrum and cerebellum, as well as in subcortical nuclei.
- The neocortex of the cerebrum is organized in six layers, each of which has a distinct functional role.
- The neuron has three principal components: the cell body; the branching dendrites, which receive signals from other neurons; and the projecting axon, which sends signals to other neurons.
- Neurons maintain electrical and chemical disequilibria in relation to the surrounding extracellular space, and many neuronal properties and processes can be understood in terms of the maintenance of these disequilibria.
- Neurons transmit information to other neurons via electrical action potentials that are propagated along axons.
- These electrical action potentials are converted to chemical signals – the release of neurotransmitter – at the synapse, the interface between the “sending” and the “receiving” neuron.
- Different types of neurons release different neurotransmitters, some of which have excitatory (depolarizing) effects on the postsynaptic neuron, some of which have inhibitory (hyperpolarizing) effects on the postsynaptic neuron.
- The net effect on a neuron receiving different types of neurotransmitter that can be released at different moments in time is that its membrane potential oscillates.
- The coordination and control of oscillations in different populations of neurons may be an important way that the flow of information in the brain is controlled.

CONTENTS

KEY THEMES

PEP TALK

GROSS ANATOMY

The cerebral cortex

THE NEURON

Electrical and chemical properties of the neuron

How neurons communicate: the action potential

How neurons communicate: the postsynaptic consequences of neurotransmitter release

Different neurotransmitters and different receptors can produce different effects on the postsynaptic cell

Housekeeping in the synaptic cleft

Neuroanatomical techniques exploit the physiology of the neuron

OSCILLATORY FLUCTUATIONS IN THE MEMBRANE POTENTIAL

Neurons are never truly “at rest”

Synchronous oscillation

COMPLICATED, AND COMPLEX

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

PEP TALK

Thinking, consciousness, decision making, language; these, and topics like them, are what you're expecting to learn about in this book. And you will. But for our explorations to be substantive, it's necessary that we establish a foundation of knowledge about how the brain is organized, structurally, and how it works. This necessitates the present chapter, which will cover neuroanatomy and cellular and circulatory neurophysiology, with little to no consideration of behavior. Thus, the following pages may be tough sledding for some readers who, for example, may have some background in psychology, but not in biology. But it will be good for you! And we just can't get around it. For example, understanding the physiology of the neuron is critical for interpreting neuroimaging data, because whereas some methods measure electrical fluctuations, others measure changes in blood oxygenation, and others measure changes in the reflectance of light. Thus, valid interpretation of data from a neuroimaging study of, say, face recognition, or emotion regulation, will depend critically on understanding what aspect of brain activity it is that is generating the signal that we are interpreting.

Later in this book, we'll see that one of the traits supported by the prefrontal cortex (PFC) is the ability to choose to engage in behavior for which the reward may not be immediate, but for which it will be worth the wait. And so, in that spirit, let's keep our eyes on the prize, and dive in!

GROSS ANATOMY

Like much of the rest of the body, most parts of the brain come in twos. There are two hemispheres, and within each hemisphere there are the five lobes of the **cerebrum** – four of them named after the four bones of the skull under which they lie (frontal, parietal, occipital, and temporal), plus the deep-lying limbic lobe, which is not adjacent to the skull (*Figure 1.2*). The brain belongs the central nervous system (often abbreviated CNS), which also comprises the spinal cord and the motor neurons that leave the cord to innervate the muscles of the body. All of these elements derive, during development of the embryo, from the neural

plate, a two-dimensional sheet of tissue that folds until its edges touch and merge to form the neural tube. In primates, which are bipedal, the CNS makes a right-angle bend between the midbrain and the forebrain such that, when standing erect, the spinal cord, brainstem, and midbrain are oriented vertically, and the **diencephalon (thalamus)** and cerebrum are oriented horizontally. Thus, for a standing human, whereas the directional terms *dorsal* and *ventral* refer to the “back” and “front” of the spinal cord, they refer to the “top” and “bottom” of the brain (*Figure 2.1*).

When you look at a brain that's been dissected out of a skull the first thing that you notice is how wrinkly it is (*Figure 1.2*, *Figure 1.4*, and *Figure 2.2*). The reason for this makes sense once you understand two facts. The first is that the computational power of the mammalian brain is in the cerebral **cortex** – the thin (roughly 2.5 mm thick) layer of tissue that covers the outside surface of the five lobes of the cerebrum – and in the cortex of two other structures, the **hippocampus** and the cerebellum. (For ease of exposition, this introductory section will focus on cerebral cortex.) The second is that the cerebral cortex is, in effect, a two-dimensional sheet, with a surface area of roughly 2500 cm². Indeed, if one could “unfold” and flatten out the cortex, it would take roughly the same surface area as a tabloid-style newspaper. (In Madison, WI, where I work and teach, I use the example of the satirical newsweekly *The Onion*, but you can pick your favorite rag.) Now we can return to the wrinkles, realizing that “crumpled” might be a better descriptor: over the course of evolution, as the computational power of the brains of vertebrate species has increased, the ever-larger cerebral cortex has needed to fold ever-tighter onto itself in order to fit inside the roughly spherical skull. The alternative, evolving a thin, 50 cm × 50 cm square (or comparably sized disk) for a skull might look okay on a creature from one of those sci-fi anime cartoons that my son watches on television, but would clearly be highly inefficient, particularly from the standpoint of the amount of “wiring” that would be needed to connect different parts of the brain together. In contrast, it has been calculated that the “crumpled up” solution with which we have evolved yields a cerebrum that is almost optimally shaped to minimize connection distances between different groups of neurons.

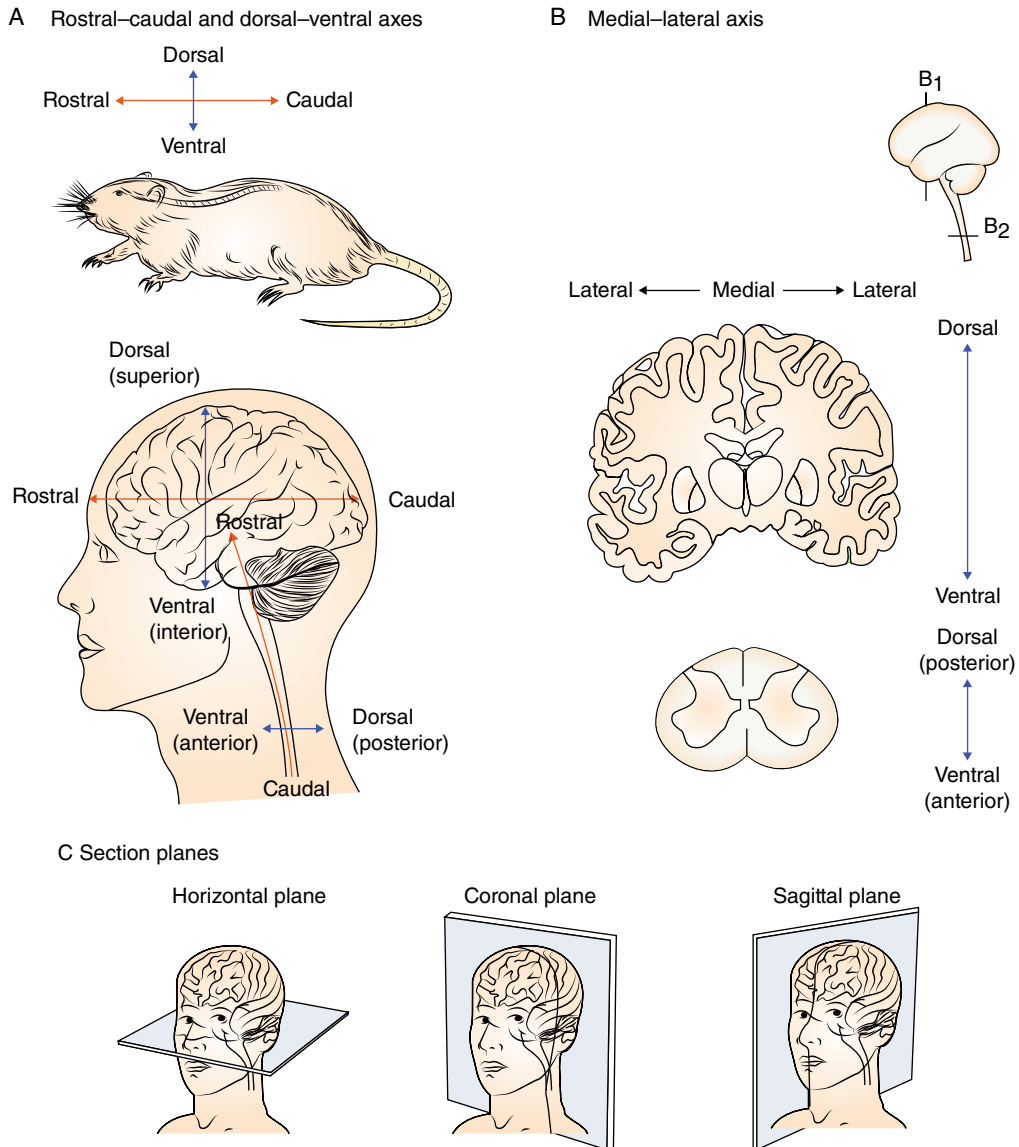


FIGURE 2.1 Terminology for anatomical orientation. **A.** The quadruped. **B.** In the bipedal human, the terminology is complicated by the right-angle bend that the neuroaxis makes between the midbrain and forebrain. **C.** The cardinal planes of section. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

When the brain is sliced, you see that the cortex looks darker (**gray matter**; although to me it looks more taupe than gray ...) than the **white matter** that takes up most of the volume of the brain (*Figure 2.2*). The gray matter is the aggregation of hundreds of

thousands of brain cells – neurons and “supporting” cells known as **glial cells**. All neurons have a branch via which they communicate with other neurons. These branches, called **axons**, conduct electrical signals that are fundamental to how neurons communicate.



FIGURE 2.2 Human brain with dorsal half of left hemisphere dissected away with a cut in the axial plane to reveal several subcortical anatomical features. Along the medial surface of the temporal lobe is the hippocampus, and the fibers projecting from its medial edge make up a tract called the fornix, which forms an arc as it courses rostrally from the caudal end of the hippocampus to terminate at the mammillary bodies of the hypothalamus. Immediately rostral to the rostral-most extent of the fornix, and to the rostral end of the hippocampus, is the anterior commissure, running in a mediolateral-oriented tract carrying transhemispheric axons between the anterior temporal lobes. Finally, the medial aspect of thalamus of the right hemisphere appears as though it is immediately ventral to the arc of the fornix. (Use the *Web 3-D Brain* to help identify these structures.) The surface of the axial cut also reveals the depth of many sulci (lined with gray matter), including the intraparietal sulcus at the caudal end of the brain. Source: Warwick and Williams, 1973. Reproduced with permission of Elsevier.

Just like the electrical power lines that carry electricity from power plants to the places where the electricity is needed, axons are insulated, so as to conduct electrical current efficiently. In the brain, this insulation is a cholesterol-laden substance called **myelin**, and the aggregation of millions of myelinated axons is what makes the white matter look white. As illustrated in *Figure 2.1*, there are three cardinal planes along which brains are sliced: in vertical cross section (“coronally”); in horizontal cross section (“axially”); or vertically along the long axis (“sagittally”). From the coronal section, in particular, as well as from *Figure 2.2*, one can

Tip 2.1

In the monkey (and sometimes in the human), the sulcus analogous to the Sylvian is often referred to as the “lateral fissure.”

appreciate many of the brain’s “wrinkles” are quite deep. The technical term for the wrinkles is **sulci**. Indeed, if you were to hold a human brain in your hand and rotate it so that you could inspect it from every angle, you’d only be able to see about one third of the cortical surface. Two thirds are “hidden” along the walls of sulci, or lining the “middle” walls of each hemisphere. In inspecting a brain like this you’d be viewing its *lateral* (i.e., “toward the outside edge”) surface. Regions nearer to the center of the volume are said to be *medial*. The “bumps” and “ridges” that are visible in the intact brain are referred to as **gyri**. The dissection featured in *Figure 2.2* provides a “window” revealing several regions of gray matter that are medial to the cortex (i.e., deep “inside” the brain). Most of these subcortical structures are nuclei – collections of cell bodies that all perform a related function. Also visible in *Figure 2.2* are large bundles of myelinated axons. The largest, the corpus callosum, carries millions of axons projecting from one hemisphere to the other.

Let’s return to the visual inspection of the intact brain that we are holding in our hands. The most obvious landmark is the midline, the gap separating the two hemispheres. From the canonical “side view” (as in *Figure 1.2*) we see the two prominent sulci that more-or-less divide the rostral from the caudal half, and we see the distinctive temporal lobe that is located ventrally and oriented parallel to the rostral-caudal axis. The sulcus that forms the dorsal boundary of the temporal lobe is the Sylvian fissure (a fissure being a really big sulcus); the sulcus that’s parallel to the caudal-most section of the Sylvian fissure, and roughly two gyri rostral to it, is the central sulcus. As its name implies, the central sulcus divides the “front half” from the “back half,” the frontal lobe from the parietal lobe. We shall see that, in functional terms, this anatomical landmark also demarcates, to a first order of approximation, the fundamental distinction of “action” (frontal) vs. “perception” (parietal, temporal, and occipital) (*Tip 2.1*). The occipital lobe, the caudal-most in the

REAL-WORLD APPLICATIONS

2.1 Myelination in development and disease

Although their authors may not know it, many laws and rules that govern modern society are influenced by one particular fact of brain development: whereas much of the circuitry of the prefrontal cortex (PFC; *Figure 1.2*) is in place by the time we reach puberty, axons responsible for transmitting signals within and from this region aren't fully myelinated until considerably later. Other neurodevelopmental processes, such as the growth of **dendrites** and development of **synapses**, also take longer to complete in PFC relative to other brain areas. This developmental lag has profound implications for behavior. The functions of the PFC won't be examined in detail until later chapters, but for the time being suffice it to say that the PFC plays a critical role in our ability to control our behavior, such as the ability to inhibit impulsive responses in favor of longer-term goals, and the ability to guide our actions according to abstract rules (e.g., laws, social conventions, long-term goals) rather than by simply responding to whatever is in front of us at a given moment in time. Thus, the developmental lag in the myelination of the PFC contributes to the fact that teenagers are statistically more likely to engage in risky behavior, and will typically be more impulsive and unpredictable than will be individuals in their mid-20s who are drawn from the same population. As a result, the developmental lag of PFC myelination can be seen as a root cause of the fact that many automobile rental companies won't rent to individuals younger than 22, and for why the age at which one can legally purchase alcohol in many parts of the United States is 21. At the other end of the developmental spectrum, the loss of myelin is a hallmark of normal aging, and can be accelerated in some neurodegenerative disorders such as Alzheimer's disease. Indeed, just as the PFC is among the last regions of the brain to fully myelinate as humans move into young adulthood, it is also among the first to begin demyelinating as they move into middle age.

Clinically, abnormal patterns of demyelination are the basis for many neurological diseases. In the CNS, the most well-known demyelinating disease may be multiple sclerosis (MS). Like many other demyelinating diseases, multiple sclerosis is the result of an abnormal immune response in which the immune system attacks the oligodendrocytes whose arms create the myelin sheath around axons. Symptoms of MS vary widely, depending on which part(s) of the CNS are affected, and can range from blurred vision (due to demyelination of axons traveling from the retina to visual nuclei of the thalamus) to movement disorders (due to demyelination of axons carrying movement-related signals in the spinal cord). MS can also impair cognition. In the **peripheral nervous system**, Guillain-Barré syndrome is also caused by an immune response that compromises the integrity of myelin. It is typified by weakness or paralysis in the extremities that migrates to the trunk. In addition to discomfort and movement problems, it can be life-threatening if it affects parts of the autonomic nervous system, such as the neurons that control breathing.

brain, sitting atop the cerebellum, has the least obvious boundaries with the two lobes with which it makes contact on its lateral surface, the temporal lobe, ventrally, and the parietal lobe, dorsally.

The tissue connecting the two hemispheres is the thick band of fibers (i.e., axons) that connect neurons of the left hemisphere to those of the right, and vice versa. This is the **corpus callosum**. If one severs the

corpus callosum, so as to separate the two hemispheres, one can then inspect the medial surface of either hemisphere (a so-called midsagittal view; *Figure 1.2.B*). This reveals the fifth lobe of the brain, the limbic lobe, which consists primarily of the anterior cingulate gyrus and the posterior cingulate gyrus, the gyri that surround the corpus callosum in its rostral and caudal aspects, respectively (*Figure 1.2.B* and *Figure 2.2*).

Very prominent in the middle of the brain, whether sliced axially, coronally, or sagittally, is a large gray-matter structure called the thalamus. Fittingly, the thalamus acts as a hub for the brain: all incoming sensory information (with the exception of olfaction) is relayed through the thalamus on its way to the cortex. Additionally, the thalamus acts as a node in several circuits, sometimes called “loops,” in which information cycles from cortex to thalamus and back to cortex again (in some cases with several other subcortical connections interposed). More detail about the anatomy and functions of these “corticothalamic” loops will be considered in subsequent chapters of the book. Because it is so intricate, and is organized anatomically according to the different brain systems with which it connects (e.g., the visual system, the auditory system, the prefrontal cortex), the thalamus is itself considered to be made up of several nuclei. Throughout the book, as we consider different systems and/or functions, we’ll also look more closely at the relevant thalamic **nucleus** (e.g., the lateral geniculate nucleus (LGN) for vision, the medial geniculate nucleus (MGN) for audition, and the mediodorsal (MD) nucleus for prefrontal cortex).

Finally, from any of the dissections that reveal the thalamus one can also see the hollow chambers that take up a considerable portion of the intracranial volume. These are the **ventricles**, the largest being the lateral ventricle. In the living organism the ventricles are filled with the **cerebrospinal fluid (CSF)** that bathes the brain on its inside (i.e., ventricular) and outside surfaces. On the outside, it circulates between a thin layer of tissue that lies on the surface of the cortex – the pia mater – and the arachnoid membrane, so named because it suggests the appearance of a spider web. Enclosing the arachnoid membrane is the rubbery dura mater, and, finally, the bone of the skull. Together, the dura, arachnoid, and pia are referred to as the meninges. (Meningitis is the infection and swelling of this tissue.) The CSF is under pressure, such that it gently inflates the brain from within the ventricles, and it is in continuous circulation through the ventricles and down the entire length of the spinal cord. For this reason, CSF extracted via a procedure called a lumbar puncture (a.k.a. a “spinal tap”), can be assayed to assess certain measures of brain health. For example, elevated levels

of certain proteins in the CSF may signal excessive cell death, as happens in Alzheimer’s disease.

The section that you have just read may well be the one that will benefit the most from *Web Link 1.1*, the 3-D brain. The book’s companion Web site also contains images from brain scans, particularly magnetic resonance (MR) images, which are the focus of *Methodology Box 2.1*, *Web Clip 2.1*, and *Web Clip 2.2*. (In addition I strongly encourage you seek out an opportunity to have the hands-on experience of holding a brain and, if possible, dissecting it. Often, especially in a formal class, the dissection will be performed with the brain of a sheep. At the level of the gross anatomy you’ll find all the structures that have just been described, and the experience will generalize to any vertebrate species (including humans) that you may subsequently study in more depth.)

The cerebral cortex

The cerebral cortex is also referred to as neocortex, in reference to the fact that it is a relatively new feature of the CNS in evolutionary terms. It is made up of six layers, which are visible when a very thin slice of cortex is viewed under a microscope. The reason that these layers look different is that each is made up of different types of cells and/or tissue that serve different functions (*Figure 2.3*).

The central layer, layer IV, is the layer that receives input from the thalamus. Thus, in the case of primary visual cortex (“area V1”), for example, the neurons of layer IV are the first cortical neurons to begin processing signals relating to, say, the photons reflecting off this page of the book. Layers II and III have subtle differences, but for the purposes of this textbook they can be thought of as being functionally equivalent, and will often be referred to as “layer II/III.” The property of layer II/III that will be emphasized here is that neurons from this layer tend to be those that send information to higher levels of cortex. As illustrated in *Figure 2.4*, it is neurons of layer II/III of the second visual processing area (“V2”) that project to a higher-level visual processing area (“V4”). (More detail on what is meant by “project to” is considered in the next section, *The Neuron*; more details on the organization of the visual system are in *Chapter 3*.) Because the flow of visual information can be construed as traveling

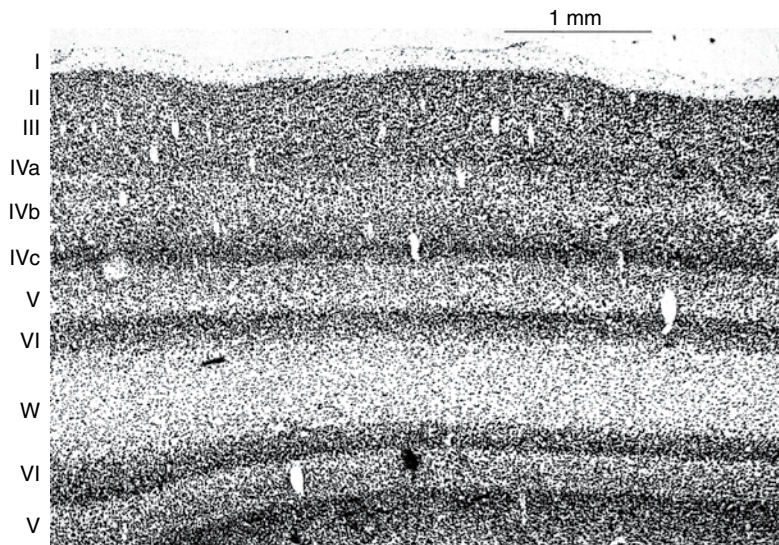


FIGURE 2.3 Cross-sectional view of cortical area V1 of the macaque, illustrating the six layers, including several subdivisions of layer IV. The tissue is from a narrow gyrus such that, at the bottom of the image, below the region of white matter (labeled “W”), one sees the deep layers of a nearby area of cortex. Source: Hubel and Wiesel, 1977. Reproduced with permission of the Royal Society.

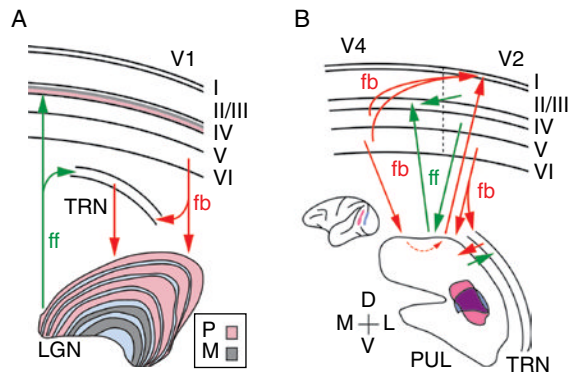


FIGURE 2.4 Schematic illustration of thalamocortical and corticocortical feedforward (ff) and feedback (fb) projections. Panel **A** illustrates the projections from the lateral geniculate nucleus (LGN; diagrammed in cross section) of the thalamus to layer IV of V1, and from layer VI of V1 back to the LGN. Note that there are six primary layers of the LGN, four containing parvocellular (P) neurons and two containing magnocellular (M) neurons (to be detailed in *Chapter 3*), and that these two types of LGN neuron project to different sublayers of cortical layer IV (to be detailed in *Chapter 5*). Also to be detailed in *Chapter 3* will be the role of the thalamic reticular nucleus (TRN), which receives collaterals from the axons carrying feedforward and feedback signals. Panel **B** illustrates the organization of feedforward and feedback pathways between two higher levels of visual cortex, areas V2 and V4 (*Tip 3.3*). “Direct” corticocortical feedforward projections go from layer II/III of V2 to layer IV of V4. “Indirect,” thalamically mediated feedforward projections go from layer V of V2 to the pulvinar nucleus of the thalamus (PUL) and then from pulvinar to layer IV of V4. Direct corticocortical feedback projections are illustrated as going from layers V and VI of V4 to layer I of V2 (the laminar specificity for this and other interareal projections varies across brain systems and across species). Indirect feedback projections go from layer VI of V4 to the pulvinar nucleus, and then from pulvinar to layer I of V2. The small diagram of a lateral view of the macaque brain relates to how these data were generated, and will be more easily understood after reading *Methodology Box 2.2*. It shows where anterograde tracer was injected into areas V4 (pink) and V2 (blue), and the diagram of a coronal slice through the pulvinar nucleus shows the overlapping projection zones from these two cortical regions. Source: Images provided by Y. Saalmann and adapted by author; see Saalmann and Kastner (2011) for more detail. Reproduced with permission of Elsevier.

Tip 2.2

For reasons unknown to this author, the layers of cortex are often written with Roman, rather than Arabic, numerals.

Tip 2.3

A generic noun for a nerve fiber, if one doesn't want to specify whether it is an axon or a dendrite, is *process*.

from the retina to the LGN of the thalamus, from thalamus to V1, from V1 to V2, and from V2 to V4, the layer II/III projections from V2 to V4 are considered feedforward connections. The neurons of layers V and VI tend to be those that send feedback projections to earlier levels of cortex. (Depending on the region of the brain, corticocortical feedback projections may arise predominantly from layer V or from layer VI.) Finally, neurons from layer VI (and sometimes also layer V) tend to be those that send their projections to the thalamus, or to other subcortical structures (*Figure 2.4; Tip 2.2*). (Note that this and the subsequent paragraph are referring to **pyramidal cells**, one of the two general classes of cortical neurons that will be described in more detail in the ensuing section, *The Neuron*.)

As we shall see in the next paragraph, it's often convenient to refer to one layer of cortex as being “above” or “below” another (as in *layer IV is located below layer II/III*). The reason that this can be problematic, however, is evident from inspecting the cartoon of a coronal slice illustrated in *Figure 2.1*. If we consider the cortex located on the ventral surface of the Sylvian fissure, and thus defining the dorsal boundary of the temporal lobe, layer II/III is, indeed, above layer IV (e.g., it is closer to the top of the page). For the cortex just dorsal to it, however, on the dorsal surface of the Sylvian fissure, the opposite is true, in that layer IV in this bit of cortex is closer to the top of the page than is layer II/III. This is a simple consequence of the fact that the “crumpled” nature of cortex means that portions of it are oriented in virtually every direction possible. A constant property of the layers of cortex, however, no matter in what part of the brain, or whether the owner of that brain is standing upright, lying on his or her back, etc., is that layer II/III is always closer to the surface of the cortex and layer IV is always closer to the white matter. Thus, their positions relative to each other are most accurately referred to as being superficial (closer to the surface) or deep (further from the surface).

And what about layer I? Layer I is distinctive from all other layers of cortex in that it contains very few cell bodies. Instead, it is made up primarily of the “antennae” (dendrites) reaching from layer II/III neurons so as to receive messages being sent by neurons in other regions (*Tip 2.3*). Indeed, this property of neurons with cell bodies located in one layer having dendrites that extend to another layer is common in all layers of cortex. Another important fact about cortical neurons is that, in addition to the properties of receiving thalamic input, sending feedforward projections, or sending feedback projections, many also send projections to other neurons located in the layer(s) above or below. As a result, groups of neurons that are “stacked” on top of each other (from layer II/III to layer VI) can often be characterized as forming **cortical columns**. As this designation implies, neurons are more strongly interconnected with other neurons within the same column than they are with neurons that are immediately adjacent but do not belong to the same column. This principle of “columnar organization” will be important for understanding the principles of visual information processing (*Chapter 3* and *Chapter 5*) and auditory and somatosensory information processing (*Chapter 4*), and may also be important in brain areas that process higher-level information (e.g., the PFC, *Chapter 14*). A final fact about the layers of cortex: depending on the region of the brain, corticocortical connections can terminate in layers I, II/III, and/or V. At this point, however, we've gone about as far as we can go without addressing the questions *What is a neuron? What are its functional units? How does it work?*

THE NEURON

The elements that give the brain its unique properties are the cells that make it up. Recall from earlier in this chapter that there are many, many different types of

cells in the brain, and that many of these are not neurons. For the time being, however, we'll focus on just two classes of cortical neurons: pyramidal cells, and interneurons. The pyramidal cell is the canonical neuron. Its large cell body has a three-dimensional shape that resembles a pyramid, with antenna-like dendrites sprouting from the cell body, particularly its

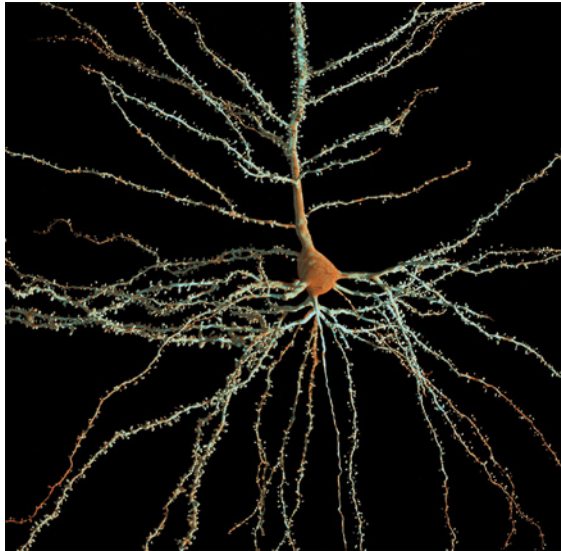


FIGURE 2.5 A pyramidal neuron from layer III of the prefrontal cortex of a monkey. Branching out from what would be the corners at the base of the pyramid-shaped soma are the basal dendrites, and projecting up toward the top of the page is the trunk of the apical dendrite, which will branch extensively when it reaches layer I (not shown). Each dendritic branch is studded with spines, submicron-sized protrusions at the end of which a synapse is formed with an axon. Extending from the base of the soma and projecting straight down and off the bottom of the image is this neuron's axon, distinguishable as the only process that is not studded with spines. To generate this image, the neuron was first microinjected with a dye, then images were acquired with a laser confocal microscope. For details about the method, see Dumitriu et al. (2011); the *Further Reading* from Morrison and Baxter (2012) demonstrates the use of this method to study the effects of aging in the prefrontal cortex. Source: Image courtesy of John H. Morrison, PhD. Reconstruction done by The Visual MD.

pointy “top,” and the axon leaving from the base of the cell (*Figure 2.5*). Typically, the dendrites, like the branches of a tree, spread themselves out over an area much larger than that taken up by the cell body itself (the cell body being the trunk in this analogy). The function of dendrites is to receive messages from other neurons (i.e., it is along dendrites that connections are formed with axons from other neurons). The axon, as we've already discussed, conveys signals from the neuron in question to other neurons. So let us now consider the nature of these electrical signals.

Electrical and chemical properties of the neuron

It may be difficult to find a single definition of “life” that any two people would find fully acceptable. Nonetheless, one (perhaps narrowly reductive) stab at a definition would be a system that is able to actively maintain itself in a high-energy state. This is a definition that draws on thermodynamics, in that it invokes the general tendency for objects in our universe to dissipate (or acquire) energy so that their own level of energy matches that of the surrounding environment. As one example, food that is removed from a hot oven and left on the countertop will cool until its temperature matches that of the kitchen. Similarly, food taken out of the freezer and left on the countertop will warm up until it achieves room temperature. In terms of their temperature, these nonliving objects passively give off or take on energy until they reach an equilibrium, which is the temperature of the kitchen. Neurons maintain life by maintaining a *disequilibrium* along two dimensions, electrical and chemical. With regard to the former, a neuron in its baseline state will maintain a negative voltage of approximately -70 millivolts (mV) relative to the extracellular space that surrounds it (see *Tip 2.4*). It achieves this by maintaining a relatively lower concentration of positively charged ions relative to the extracellular space. In addition to

Tip 2.4

Note that the resting potential of -70 mV refers to a generic neuron. Different cell types will have different resting potentials.

this electrical disequilibrium, the neuron at rest maintains a chemical disequilibrium by “hoarding” a high concentration of potassium (K^+) ions relative to the extracellular space, and maintaining a relatively low concentration of sodium ions (Na^+) relative to extracellular space. (Recall that the membranes of all cells in the body are made up of a lipid bilayer that, because

it repels water, separates the watery cytoplasm (and all dissolved in it) inside the cell from the watery extracellular fluid (and all dissolved in it).)

Hey, did someone say “watery extracellular fluid”? That’s a cue to discuss **diffusion tensor imaging** (DTI), a tool for in vivo neuroanatomy research (see *Methodology Box 2.1*).

METHODOLOGY BOX

2.1 MRI and DTI – methods for in vivo neuroanatomy

T1-weighted MRI Nuclear magnetic resonance (NMR) refers to the fact that the *nucleus* of an atom (a proton, in the case of hydrogen), *resonates* at a preferred frequency when in the presence of a strong *magnetic* field. For our purposes, we can understand this as owing to the fact that an “unpaired” nucleus (i.e., no neutron paired up with the H atom’s lonely proton) rotates. In the jargon of nuclear physics, this rotation is a “spin.” NMR imaging, now commonly referred to as magnetic resonance imaging (MRI, see *Tip. 2.5*), is based on the fact that the application of a radiofrequency (RF) pulse at just the right frequency – the resonant frequency of the H atom – can perturb the axis around which the proton is spinning. Because the atom is in the presence of a strong magnetic field produced by the MRI scanner (the “ B_0 ” field, pronounced “bee naught”), the low-energy state for that atom is for the axis of its “spin” to be aligned with that magnetic field. (In MRI, this is the field whose strength is measured in units of tesla, resulting in scanners being labeled, for example, “1.5 tesla” (T), or “3T” scanners.)

Now, where were we ... Oh, yes ... In the process of returning from the higher-energy state into which it has been pushed by the RF pulse, to the low-energy state of alignment with the B_0 field (a process referred to as “relaxation”), the H atom gives off energy. It is this energy that creates the MR image. Different ways of delivering the perturbing RF pulse, referred to as pulse sequences, can influence nuclear spins in different ways, and thereby emphasize different types of tissue. Additionally, different ways of measuring the relaxation signal can yield different kinds of information about the tissue being imaged. One component of this signal, the T1 component, is particularly well suited for distinguishing white matter from gray. This is illustrated in *Figure 2.6*, which features a “T1-weighted” image of a virtual slice through a brain. T1-weighted images (“T1 images,” or, sometimes, “T1s” for short) are the most common structural MR images that one encounters in the cognitive neuroscience literature. (More detail and examples in *Web Clip 2.1* and *Web Clip 6.2*.)

DTI tractography Here’s where the watery extracellular fluid comes in. Within tracts of myelinated axons there are teeny gaps between each axon. As a consequence, water molecules travel much more quickly along such a bundle than across it. This motion, or *diffusion*, of water molecules can be measured with a particular kind of pulse sequence. The resultant data take the form of a 3-D volume (called a *tensor*) that indexes how symmetrically, or otherwise, water molecules in a particular location within the brain are diffusing. A perfectly spherical tensor, said to be “isotropic,” would indicate that water is diffusing freely in all directions. Fiber tracts, in contrast, produce “anisotropic” tensors shaped more like a rugby football. By estimating the most probable route that a series of such tiny end-to-end “footballs” makes through a

Methodology Box 2.1 Continued

section of brain, one can infer the existence of a fiber tract, as illustrated in *Figure 2.6*. (More detail and examples in *Web Clip 2.2*.)

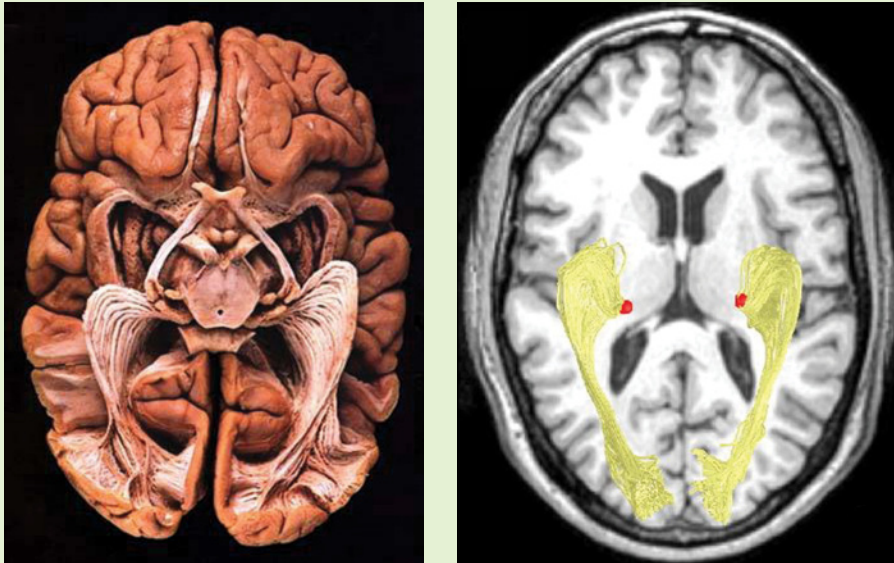


FIGURE 2.6 Image on left shows a human brain dissected to reveal the pathways from retina, via LGN, to V1. Particularly striking are the elegant optic radiations, the myelinated fibers running between the thalamus and the primary visual cortex. To expose these pathways and structures, the anatomist dissected away tissue from the ventral surface of the brain, up to and including the ventral bank of the calcarine fissure. Thus, only the dorsal bank of the calcarine and the fibers of the optic radiations that project to it are seen here. (Some of this will make better sense after having read the first section of *Chapter 3*.) The image on the right shows a slice from a T1-weighted anatomical MR scan (grayscale) oriented to include both the LGN (pseudocolored red) and the calcarine sulcus. Superimposed, in yellow, is an image of the path of the optic radiations as estimated via MR-derived diffusion tensor imaging (DTI) tractography (see *Web Clip 2.1* and *Web Clip 2.2*). Source: Wandell and Winawer, 2011. Reproduced with permission of Elsevier.

Tip 2.5

There is nothing radioactive about NMR. However, because the word “nuclear” can provoke strong reactions (particularly in someone about to undergo a medical imaging procedure!), the alternative “MRI” has become the more common term in the context of medical imaging.

The consequence of the electrical and chemical imbalances in the neuron is that if a hole were to be poked in the membrane of a neuron, two events would happen: (1) the electrical gradient would be reduced by an influx of positively charged ions into the cell (including Na^+ ions); (2) simultaneously, the chemical gradient would be reduced by an efflux of K^+ from the high- K^+ concentration inside the cell to the low- K^+ concentration outside the cell. (And note

Tip 2.6

Because the resting potential of the neuronal membrane is negative, the convention is to say that a current whose effect is to move this potential closer to 0 mV is a *depolarizing* current, and a current whose effect is to move the potential further from 0 (i.e., making it more negative) is a *hyperpolarizing* current.

that the influx of Na^+ ions from (1) would also be reducing the chemical gradient.) As you might imagine, however, the workings of the neuron are much more orderly and regulated than the poking of holes. Indeed, the membrane of the neuron is punctuated with channels that can open and close, thereby allowing the passage of certain types of ions. Most of these channels have differential permeability to different types of ions, and are only activated under certain conditions. One example is the voltage-gated Na^+ channels that are concentrated at (or near, depending on the type of neuron) the **axon hillock**, the region where the bulky cell body funnels down into the long tube that is the axon. These channels are closed when the neuron is at rest (i.e., with a membrane potential of -70 mV). However, when the neuron has been sufficiently **depolarized** (e.g., the voltage raised to roughly -20 mV (again, this varies depending on cell type); see *Tip 2.6*), these channels open, and allow for the very rapid influx of Na^+ ions. Recall that this influx is propelled by both the electrical gradient and the chemical gradient that exists across the cell membrane. This influx of Na^+ ions gives rise to the action potential (or “spike” in neurophysiology vernacular), which is the basis for neuronal communication.

How neurons communicate: the action potential

A neuron initiates an action potential by taking advantage of the local depolarization produced by the influx of Na^+ ions that occurs at the axon hillock. It has to “act fast,” because although this event produces a short-lived “cloud” of depolarization in the vicinity of the axon hillock, this “cloud” will quickly dissipate due to the rapid diffusion of Na^+ ions throughout the intracellular space (recall that there’s a

low concentration of Na^+ ions inside the cell). The neuron exploits this temporary depolarization by having another cluster of voltage-gated Na^+ channels that is “strategically” located just a little way along the axon. The channels of this cluster experience the brief pulse of depolarization from the hillock and they, in turn, open. Next, guess what? There’s *another* cluster of voltage-gated Na^+ channels even further along the axon ... This pattern repeats, and in this way an action potential is propagated along the length of the axon until it reaches the dendrites of the next neuron. (You may be wondering, “But how did the initial depolarization that triggered the activity at the axon hillock come to pass?” Stick with me, and we’ll come to that.)

When the action potential arrives in the vicinity of a dendrite from another neuron, it encounters a change in the morphology (i.e., shape) of the axon. This is a swelling of the axon into what is referred to as a synaptic bouton, or presynaptic terminal (*Figure 2.7*). For a neural signal to pass from one cell to another, there needs to be a synapse linking them. At the presynaptic terminal one finds the machinery necessary for converting the electrical signal of the action potential into the chemical signal that will be transmitted to the postsynaptic terminal. (Note from *Figure 2.7* that the pre- and postsynaptic cells do not touch, there is a gap (the **synaptic cleft**) between them.) That is, the electrical signal does *not* jump to the next neuron. Instead, the rapid depolarization triggered by the action potential opens voltage-gated calcium-ion (Ca^{++}) channels located in the membrane of the presynaptic terminal. Not unlike Na^+ , Ca^{++} ions are in much higher concentration outside than inside the neuron, and thus the opening of voltage-gated Ca^{++} channels results in an influx of Ca^{++} ions into the presynaptic terminal. Ca^{++} ions have a special status in the nervous system, in that they can initiate many biochemical processes within the cell. At the presynaptic terminal, the influx of Ca^{++} ions triggers a cascade of events that results in the release of neurotransmitter into the synaptic cleft (*Figure 2.7*; see *Tip 2.7*). (Note that it is almost never the case that a neuron makes contact with just one other neuron. Rather, axons have multiple branches (in some cases thousands), each of which ends in a synaptic terminal. The action potential propagates equally along each branch.)

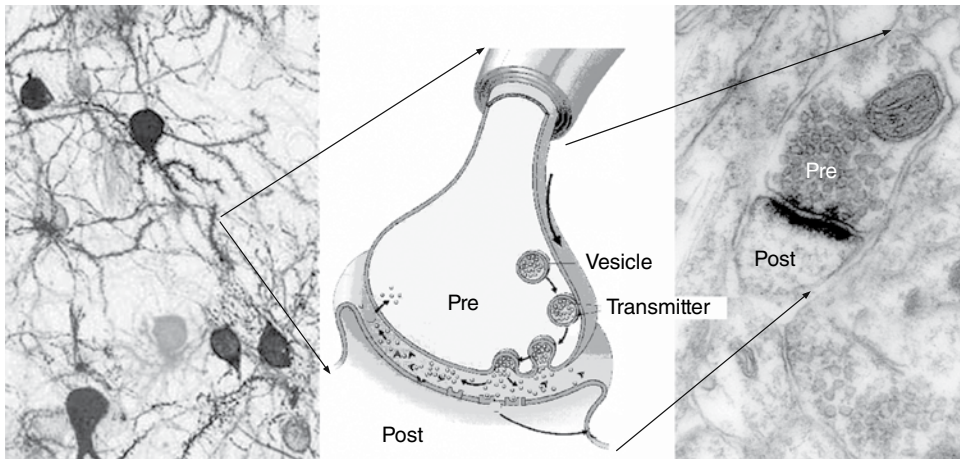


FIGURE 2.7 Left-side panel: light microscope image of neural tissue with cell bodies and processes visible. Many of the dendrites are studded with dendritic spines, including the one from which a tiny region is blown up to illustrate a synapse in the center panel. Center panel: diagram of a synapse, illustrating (with arrows) the events triggered by the arrival of an action potential: vesicles filled with neurotransmitter in the presynaptic terminal (“pre”) are transported to the membrane, where they fuse with it to release neurotransmitter into the synaptic cleft; subsequently molecules of neurotransmitter (or, depending on the type of synapse, a catabolyte), are reuptaken back into the presynaptic terminal. Right-side panel: an electron microscope image of a synapse, with vesicles visible in the presynaptic terminal, and the “postsynaptic density,” partly comprised of membrane-spanning receptors, visible as a dark band at the postsynaptic terminal. Source: Buszáki, 2006. Reproduced with permission of Oxford University Press, USA.

Tip 2.7

The electric-to-chemical-to-electric conversion is typical of most synapses in the CNS. However, some classes of neurons in the brain are “electronically coupled,” and can pass current directly between themselves.

How neurons communicate: the postsynaptic consequences of neurotransmitter release

On the postsynaptic side of the synapse, the membrane is studded with *neurotransmitter receptors*, each of which can bind with a molecule of the just-released neurotransmitter. This, in turn, will, depending on the type of receptor, produce one of any number of possible effects in the postsynaptic neuron. The most straightforward of these to describe, and one that’s satisfactory for our present purposes, is to open a **ligand-gated** Na^+ channel – that is, a molecule of

neurotransmitter acts like a key by fitting into a special slot in the receptor and “unlocking” it. With its channel now open, the receptor allows the influx of Na^+ ions into the postsynaptic terminal of the dendrite (See *Thought Box 2.1*). Thus, in our example, the binding of a molecule of neurotransmitter to a receptor has the effect of generating a small “cloud” of depolarization in the vicinity of this receptor within the postsynaptic terminal. Now if only one such event happens, the cloud of depolarization quickly dissipates. If, however, many molecules of neurotransmitter have bound with many receptors, particularly if this has happened more-or-less simultaneously at many synapses (i.e., if this neuron is on the receiving end of a volley of many action potentials arriving simultaneously from many different presynaptic neurons) then all these individual pulses of depolarization may combine, via their summation, to the point at which the cell’s membrane voltage is raised to the neighborhood of -20 mV , in which case an action potential may be triggered in this postsynaptic neuron. (A mechanism

THOUGHT BOX

2.1 How do membrane channels work?

The membrane-spanning channels of neurons are made up of complexes of proteins that span the membrane from intra- to extracellular space. The channel's walls are formed by several subunits, as is illustrated in *Figure 2.8*. A channel's specificity for one type of ion is determined by the physical size of its pore

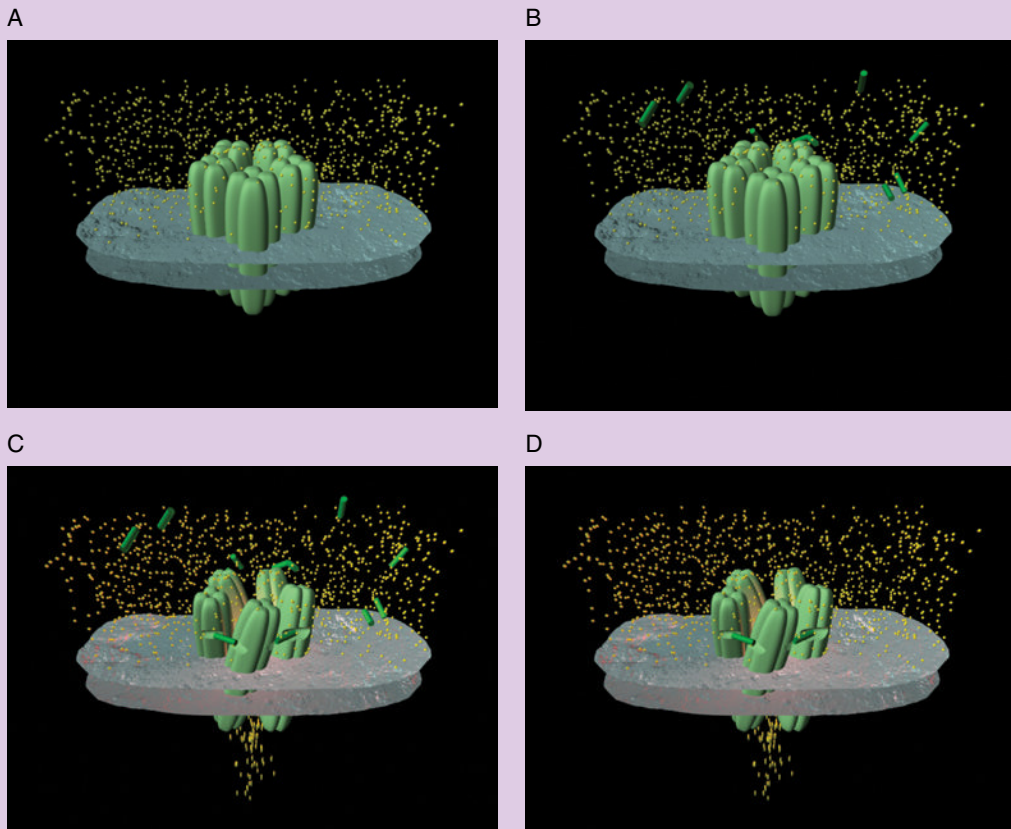


FIGURE 2.8 Four panels illustrating the effects of the neurotransmitter GABA binding with a membrane-spanning GABA_A receptor (a subtype of GABA receptor). **A.** The receptor is composed of five separate proteins, each one consisting of four membrane-spanning domains (green “cigars”). As a functional unit, the receptor spans the neuronal membrane (gray planes), and provides a pathway for the inflow of ions (yellow dots represent a cloud Cl⁻ ions in the extracellular space). The receptor is shown in its resting, closed state, with no neurotransmitter bound, and no available path for ions to flow across the membrane. **B.** Shorter, darker green cylinders represent GABA molecules having been released into the synaptic cleft. (I.e., an action potential in a GABAergic (*Tip 2.11*) neuron has triggered the release of GABA from the presynaptic terminal (not shown) into the cleft.) **C.** Binding of GABA molecules at two binding sites results in conformational change of receptor proteins (illustrated here as a mechanical twisting of subunits) which opens the pore of the receptor, thereby allowing Cl⁻ ions to enter the neuron. **D.** For some period of time after unbound GABA has been cleared from the synapse, the receptor remains active. Source: Images courtesy of Mathew V. Jones, University of Wisconsin-Madison. Reproduced with permission.

Thought Box 2.1 *Continued*

(e.g., Na^+ channels are simply too small to allow large ions or molecules to pass through), as well as by the properties of binding sites located in the pore of some channels that result in the preferential permeability to some ions over others. When closed (i.e., “at rest”), the pore is too narrow for anything to pass through. When activated, the channel twists (due to a change in conformational state of one or more of the subunits), thereby enlarging the pore and allowing ions to pass through. Depending on the type of channel, activation is accomplished by the binding of a ligand to the binding site (as is the case with AMPA and GABA receptors), by a change of membrane voltage, or by activation of the channel via chemical signaling from within the cell. The latter is typical of many channels that are influenced by a neurotransmitter other than glutamate or GABA. In the case of the neurotransmitter dopamine, for example, the dopamine receptor is not part of a ligand-gated channel. Instead, the binding of dopamine to a receptor triggers a cascade of biochemical events inside the postsynaptic terminal, and a downstream consequence of this can be the opening of a membrane-spanning channel. Note that such a multi-step cascade is likely to take longer to influence the voltage of a neuron than is a ligand-gated channel. This is one of the reasons that neurotransmitters such as dopamine are considered to be neuromodulators – their effect is to modulate the responsivity of the neuron to the faster-acting, “driving” effects of glutamate and of GABA.

by which the action potentials from many neurons can be coordinated to all arrive at a postsynaptic neuron at the same time is considered near the end of this chapter, in the section *Synchronous oscillation*.)

Different neurotransmitters and different receptors can produce different effects on the postsynaptic cell

The postsynaptic effects of neurotransmission that were described in the previous paragraph apply to the **excitatory neurotransmitter** glutamate binding to a particular kind of glutamate receptor called the AMPA receptor (see *Tip 2.8*). There are other glutamate receptors in the brain that work differently, and we will consider them in some detail when we discuss memory functions. **Glutamate** is released by cortical pyramidal cells, and because these are the most numerous neurons found in the cortex, glutamate is the predominant excitatory neurotransmitter in the CNS. Equally important for the functioning of the brain is **inhibitory neurotransmission**, and this is most commonly carried out by the neurotransmitter **GABA** (see *Tip 2.9*). GABA is the primary neurotransmitter released by **inhibitory interneurons**, neurons whose primary function is to **hyperpolarize** neurons (i.e., to make their voltage more negative). The

Tip 2.8

Neurotransmitter receptors are often named after chemicals that can mimic the effects of their naturally occurring ligands. Thus, the AMPA receptor is named after the chemical that binds this type of glutamate receptor, but not other types of glutamate receptor: α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid.

Tip 2.9

GABA is the acronym for the molecule gamma-aminobutyric acid.

most common GABA receptor produces its hyperpolarizing effects by opening channels that pass chloride ions (Cl^-), and because the extracellular concentration of Cl^- ions is considerably higher than the intracellular concentration, opening these channels results in an influx of negatively charged Cl^- ions, and, thus, a negative shift in the membrane potential (*Figure 2.8*). The reason that this effect on the postsynaptic neuron is inhibitory is because it pulls the

Tip 2.10

Positively charged ions, as a category, can be referred to as **cations**, negatively charged ions as **anions**. These terms are often used when describing ion channels that are not specific to one type of ion.

Tip 2.11

The suffix used to convert the names of many neurotransmitters from nouns to adjectives is “ergic.” Thus, for example, a glutamate-releasing synapse is called “glutamatergic,” a GABA-releasing neuron “GABAergic,” and so on.

membrane voltage away from the trigger potential of -20 mV, thereby inhibiting the neuron’s probability of firing an action potential (*Tip 2.10*).

Housekeeping in the synaptic cleft

Although the pre- and postsynaptic terminals of a synapse are very close together, there is no guarantee that any one molecule of neurotransmitter will bind with a receptor. Rather, binding is a probabilistic event, with the probability of binding influenced by many factors, including the amount of neurotransmitter released by an action potential, and the density of receptors in the postsynaptic membrane. It is also important to know that those molecules of neurotransmitter that do bind with a postsynaptic receptor do not enter the postsynaptic cell. Rather, after some period of time, they disengage from the receptor and float back into the synaptic cleft. The nervous system thus has to solve the problem of clearing out the synapse after each action potential, lest the lingering molecules of neurotransmitter in the synapse continue to influence the postsynaptic neuron for longer than intended. The nervous system has evolved many different solutions to this problem that are implemented at different synapses. Common to many is the process of **reuptake**, whereby molecules of neurotransmitter are taken back into the presynaptic terminal via specialized complexes of membrane-spanning

proteins. Reuptake is an active process that requires energy, because it entails moving molecules of neurotransmitter from a low-concentration environment (the extracellular space) into a high-concentration environment (the presynaptic terminal). That is, it entails working against a chemical gradient (see *Thought Box 2.2*). At glutamate synapses, the reuptake occurs both at the presynaptic terminal and at astrocytes that are located adjacent to the synapse. (In astrocytes, glutamate is converted into a precursor metabolite, which is transferred back to the presynaptic pyramidal cell for conversion back into glutamate.) At GABA synapses, the neurotransmitter is reuptaken into the presynaptic terminal by the GABA transporter. At synapses for many of the neuromodulatory transmitters, there are enzymes that either break down the molecules of neurotransmitter into smaller, constituent molecules, or transform them into a different molecule, the result in either case being a product that can no longer bind with the postsynaptic receptors. These now-inert molecules are then taken back into the presynaptic neuron, where they are reconverted into molecules of neurotransmitter. There are entire textbooks devoted to the details of how neurotransmitters are synthesized and metabolized (often falling under the rubric of “neuropharmacology”), and others to the details of the trafficking, release, and reuptake of neurotransmitters at the presynaptic bouton. Although we won’t devote much more verbiage in this book to the detailed workings of the synapse, the facts that we have covered up to this point are important for understanding two general properties that are highly relevant for cognitive neuroscience. One, introduced in *Thought Box 2.2*, is how the metabolic demands of the synapse trigger physiological signals that are the basis of neuroimaging signals that can be used to study the brain. The second, detailed in *Real-World Applications 2.2*, is that the synapse is the target of virtually every drug (whether medicinal or recreational) that affects cognitive functioning. (Finally, I recommend a visit to *Web Link 2.1*, which illustrates how an understanding of the neuron at this level of detail is being used by computer scientists to build more efficient computers, as well as more accurate computer simulations of the brain itself.)

THOUGHT BOX

2.2 The metabolic demands of the action potential and synaptic activity

The generation of an action potential entails reversing the electrical and chemical imbalances that are characteristic of the neuron at rest. How the neuron restores these imbalances turns out to be the basis for many neuroimaging techniques. One key mechanism is the so-called **Na⁺/K⁺ pump**, a membrane-spanning complex of proteins that act as a “reverse channel.” At the cost of one molecule of adenosine triphosphate (ATP) per cycle, the Na⁺/K⁺ pump expels three Na⁺ ions and draws in two K⁺ ions. Also highly energy-demanding is the process of releasing molecules of neurotransmitter into the synaptic cleft (exocytosis). ATP, the “fuel” that is needed for these operations, is derived from the processing of glucose in mitochondria inside the neuron. Glucose, in turn, is delivered to cells via the blood stream. Herein lies the linkage to the signal that is measured by **functional magnetic resonance imaging (fMRI)** and some applications of positron emission tomography (PET), both to be discussed in more detail in *Chapter 6* and subsequent chapters), as well as by certain types of functional tissue imaging (*Chapter 3*): elevated activity in neurons creates a need for arterial blood, and the resultant increase of oxygenated blood, or of glucose, to active parts of the brain is what these methods detect. (The principle is the same for all tissues of the body. Think about whether more blood flows to the muscles of your arm when you are lifting something heavy, or when your arm is just hanging by your side.)

Interestingly, neurotransmitter reuptake, although a metabolically expensive process, isn’t directly responsible for the consumption of ATP. Instead, it leverages electrochemical ion gradients across the membrane via a process called **cotransport**. A useful analogy is to consider a revolving door. To get it “inside,” a molecule of neurotransmitter is yoked to a Na⁺ ion. Now we have a situation in which the chemical gradients of the yoked components are in opposition – the low intracellular concentration of Na⁺ provides an inward “pull,” and the opposite is true for the neurotransmitter. It’s as though two people have entered a door that’s rotating counterclockwise (anticlockwise), and one (the “Na⁺ ion”) is pushing the door in the counterclockwise direction so as to enter, while the other (the “molecule of neurotransmitter”) is pushing in the opposite way. (The role of the neurotransmitter can be played by a two-year-old in our analogy: *No, Mommy, I don’t want to go into that store!*) The stalemate is broken by adding a third “person” who is inside and who wants to leave: cotransport also entails the simultaneous transport of a molecule of K⁺ along its concentration gradient, from inside the cell to outside. “Two against one,” the molecule of neurotransmitter “loses,” and it ends up back inside the neuron.

If cotransport “passively” takes advantage of chemical gradients, why is it metabolically expensive? The answer is that although the reuptake process itself doesn’t consume ATP, its operation decreases the resting-state imbalance of higher intracellular concentration of K⁺ and higher extracellular concentration of Na⁺. Thus, each cycle of the reuptake processes prompts a compensatory, energy-expending cycle of the Na⁺/K⁺ pump to in order to restore the imbalance.

REAL-WORLD APPLICATIONS

2.2 Most drugs that affect the brain have their effects at the synapse

Whether one considers pharmaceutical medications that are prescribed to treat psychological, psychiatric, or neurological disorders, or “recreational” drugs that people take to get high, the synapse is far and away the site of action of most neuroactive drugs. Many recreational drugs have their effects by mimicking endogenous ligands. Thus, for example, the active component of heroin mimics endogenous compounds called opioids and binds to opioid receptors. The active component of marijuana mimics endogenous compounds called cannabinoids and binds to cannabinoid receptors. Cocaine, in contrast, has its effect by blocking the reuptake of dopamine into the presynaptic terminal, thereby keeping molecules of dopamine in the synapse longer, and exaggerating the effect of dopamine release. Interestingly, the mechanism of the class of antidepressant medications known as selective serotonin reuptake inhibitors (SSRIs; examples include Prozac and Zoloft) work on the same principle, but on a different neurotransmitter system.

Some drugs don’t directly gate receptors, but, instead, modulate their action. Benzodiazepines (such as diazepam (first marketed as Valium)), for example, bind to a site on the GABA_A receptor, and have the effect of increasing its Cl⁻ conductance when GABA is released. (That is, benzodiazepines don’t directly activate these receptors, but, when bound to the receptors, they potentiate the effects of GABAergic neurotransmission. See *Figure 2.8*.)

Finally, many naturally occurring *neurotoxins* also do their work at the synapse. For example, curare is a toxin that indigenous peoples of Central and South American rain forests have learned to extract from the skin of brightly colored “poison dart frogs.” Curare paralyzes its victims by blocking the release of the neurotransmitter acetylcholine (ACh) at the neuromuscular junction, the synapse at which nerve endings instruct muscle cells to contract. (Diabolically, curare has minimal effects on the CNS, and so paralyzed victims are presumably aware of their condition as the hunters close in)

One exception to this synaptocentric pattern is tetrodotoxin (TTX), the deadly poison administered by puffer fish, blue-ringed octopi, and many other sea creatures that one should avoid. Rather than working at the synapse, TTX interferes with the ability to generate action potentials by blocking the voltage-gated Na⁺ channel.

Neuroanatomical techniques exploit the physiology of the neuron

Now that we have an understanding of some aspects of how the neuron “works,” we can return to a topic from earlier in this chapter – the **laminar** organization of cortical connectivity – to consider how this knowledge has been acquired. Let’s start by considering that no matter how long, or multiply branched, is an axon, it’s part of one cell. Thus, there needs to be a mechanism, for example, for the transport of proteins that are synthesized from DNA in the nucleus of the

neuron to synapses at far reaches of each axonal branch. (For me, the “gee whiz” factor is particularly high when I think of the giraffe, or of massive whales) And so, whereas our attention up to this point has been on various types of trans-membrane transport, intracellular mechanisms for **axonal transport** are critical for the healthy functioning of the neuron. The reason for bringing this up now is that, for centuries (whether they have known it or not), anatomists have taken advantage of axonal transport to explore connectivity in the brain (*Methodology Box 2.2; Figure 2.9*).

METHODOLOGY BOX

2.2 Cell biological methods for measuring neuronal connectivity

Axons are microscopically small. Although bundled tracts of thousands of axons can be visible to the naked eye (e.g., the optic nerve, tract, and radiations; *Figure 2.6*), this sheds no light on the destination of individual axons. One strategy for determining precisely where axons project to is to inject a substance into a region of the brain, let the natural processes of intracellular transport “do their thing” (i.e., transport the substance from the region of uptake to other parts of the neuron), and later look for where the substance has been transported. Over the past 100 years or so, sometimes through trial-and-error, sometimes as a result of sophisticated understanding of cell biology, anatomists have devised many ways to address different questions about neuroanatomy. For our purposes we can think of three categories of cell-based (as opposed to neuroimaging-based) neuroanatomy: anterograde transport; whole-cell staining; and retrograde transport.

Anterograde transport would be used, for example, to determine whether, and how, neurons from somatosensory area S1 project to somatosensory area S2. To do the experiment, one would inject tracer into S1, wait for a set period of time, then sacrifice the animal, slice the brain, process the tissue, and, finally, inspect the tissue under a microscope to observe where the tracer ended up. Such an experiment depends on several factors. First, the tracer needs to be taken up by the dendrites and/or cell bodies of the neurons of interest. Second, one needs to know the amount of time it takes for the tracer to be transported to the ends of the neuron. Third, there needs to be a reliable way to detect the tracer under a microscope. With regard to the passage of time, this becomes particularly important if the tracer can “jump” the synaptic cleft, and thereby be passed from one neuron to another. Exquisitely detailed studies have been performed with such tracers such that, when different animals are sacrificed at different times relative to the injection, the anatomist can determine the monosynaptic connectivity of a region, its disynaptic connectivity, and so on. With regard to detection, some tracers are visible in and of themselves, others require the tissue to be stained or otherwise prepared so as to make the tracer visible.

Whole-cell staining is just what it sounds like. A nice demonstration is seen in the left-hand panel of *Figure 2.7*. The trick with this method is to use a stain that is selective for a particular kind of cell. For example, we’ll see in *Chapter 4* that to study the local architecture of pyramidal vs. interneurons in primary somatosensory cortex, the researchers take advantage of the fact that one class of interneurons contains the protein parvalbumin, but other interneurons and pyramidal cells do not.

Retrograde transport is the opposite of anterograde transport. It relies on a tracer that is taken up at the presynaptic terminal (perhaps by “hijacking” the cotransport mechanism), then carried “backwards” along the axon to the cell body and dendrites. *Figure 2.9* illustrates how this approach can be used to determine the laminar organization of projections from S1 to S2 vs. to the nuclei of the “somatosensory thalamus” (see *Tip 2.12*).

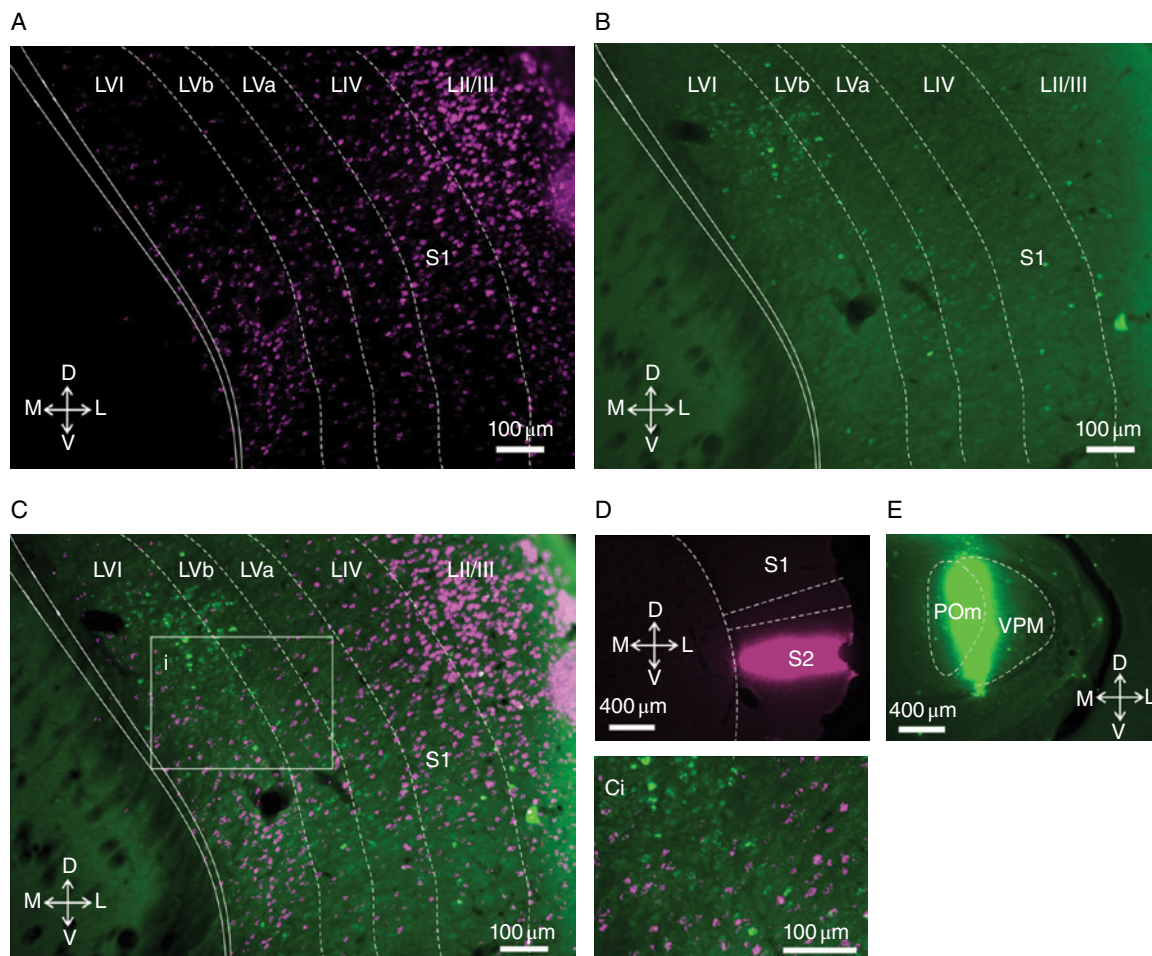


FIGURE 2.9 A neuroanatomical tracer study of the laminar organization of projections from primary somatosensory cortex (S1). To follow the logic of the experiment, look first at panel **D** – this shows the region of secondary somatosensory cortex (S2) where the (purple-colored) tracer was injected. It is a fluorescent, retrograde tracer called “Retrobeads.” Panel **A** shows the distribution of labeling resulting from the injection shown in **D**. That is, the purple Retrobeads were taken up at presynaptic terminals in S2 (panel **D**) and retrogradely transported back to the cell bodies (panel **A**) of the neurons whose axons formed those presynaptic terminals (that is, of the neurons whose axons projected to S2). Note that the labeling is heaviest in layer II/III of S1, meaning that it is neurons in this layer that send the heaviest feedforward projection from S1 to S2. (Note, also, that this laminar organization is not absolute – biological systems are rarely as tidy and lawful as one might want them to be.) Next, let’s go to panel **E**, which illustrates the injection site of green Retrobeads into the posterior medial nucleus (POm) and ventral posterior posterior medial (VPM) nucleus of the thalamus. The green spots concentrated in the upper left-hand corner of panel **B** are the retrograde labeling resulting from this thalamic injection. This indicates that the S1 neurons projecting to thalamus are located primarily in layers VI and Vb. Panel **C** illustrates an overlay of **A** and **B**, to emphasize the distinct projection profiles of the superficial vs. the deep layers of S1. Panel **Ci**, a blow-up of the box in panel **C**, shows the absence of “double-labeling,” the presence of which would have meant that individual neurons project to both S2 and the thalamus. (Addressing this question was, in fact, the primary reason that this experiment was performed.) The key in the lower-left corner of each image indicates the dorsoventral and mediolateral axes. Source: Petrof, Viaene, and Sherman, 2012. Reproduced with permission of Wiley.

Tip 2.12

Note that I am deliberately choosing not to name various neuroanatomical tracers, because there are hundreds, and our purpose in this chapter is to convey an idea of how neuroanatomical connectivity is determined, not to teach the details of how one would perform such an experiment.

OSCILLATORY FLUCTUATIONS IN THE MEMBRANE POTENTIAL

Neurons are never truly “at rest”

As previously noted, the small cloud of depolarization produced by the opening of any single ligand-gated Na^+ channel typically won't, by itself, be enough to trigger an action potential, particularly when the channel is located on a distal dendritic branch, far from the cell body and its axon hillock. If this one channel were to periodically open then close, open then close, what one would observe at the cell membrane in the immediate vicinity of the receptor would be a slight depolarization – say, from -70mV to -69mV – followed by a return from -69mV back to -70mV , followed by another rise back to -69mV , and so on. That is, the local membrane voltage would oscillate. Now, of course, we already know that the factors influencing the membrane voltage are much more complicated than the behavior of a single ligand-gated ion channel. For starters, a single synaptic event (i.e., an action potential in the presynaptic neuron) is likely to activate hundreds, if not thousands, of individual channels. These individual depolarizing events will add together (they will “summate”). So, too, will the depolarizing effects produced by activity at other synapses. As a result, the overall effect that is experienced by the synaptic membrane will depend on the synchrony of incoming action potentials from glutamate-releasing axons. That is, if many action potentials arrive at about the same time, the aggregated depolarizing effect will be much larger than if the same number of incoming action potentials are spread out in time, because the depolarizing effect of the earliest-arriving action potentials will

have started to dissipate (also referred to as “decay”) before the depolarizing effect of the later-arriving ones starts to kick in. A second factor that influences fluctuations in overall membrane potential is the hyperpolarizing effects of action potentials at GABAergic synapses. What emerges from this cacophony of thousands of events at individual channels, some depolarizing and some hyperpolarizing, is an ongoing pattern of fluctuations in the dendritic membrane potential that can be measured with an electrode in the extracellular space. Because the dendritic branches from many, many neurons are located in close proximity to each other, it is typically not possible to isolate the fluctuations attributable to a single neuron. Rather, what is recorded is the **local field potential (LFP)**, a reference to the fact that it's the local region of cortex where these dendrites are located (i.e., the “dendritic field”) from which the signal is being measured (see *Methodology Box 2.3*; *Figure 2.10*; and *Web Clip 2.3*).

Despite the fact that there can be literally hundreds of thousands of individual channels distributed across the tens to hundreds of neurons contributing to the LFP recorded by any one electrode, the LFP is nonetheless often observed to fluctuate in an orderly oscillation. Indeed, different oscillatory frequencies, or “states,” are so typical of the cortex that they've been given names. One common state is when the LFP oscillates with roughly 40 up–down cycles per second (i.e., 40 Hz); the range of oscillatory frequencies within which this falls, from roughly 30 Hz to roughly 80 Hz, is referred to as the **gamma frequency band** (or, “the gamma band,” or, sometimes, in the shorthand slang of the laboratory, just “gamma”). (This and other frequency bands are illustrated in *Figure 4.3*.) How can the brain orchestrate what we've just referred to as a cacophony of individual single-channel events into a steady, predictable up-and-down rhythm? One important factor is that neurons are wired up in a way that facilitates the control of their oscillations. For many pyramidal cells, for example, inhibitory inputs from interneurons tend to synapse near the base of the dendrite, or even on the cell body itself. This positioning allows these inhibitory synapses to act as a kind of “gate” that either does or does not allow the depolarizing currents from more distally located glutamatergic synapses to

METHODOLOGY BOX

2.3 Extracellular recording: action potentials and LFPs

Although many methods have been devised to study the electrical properties of neurons, most have made use of some variant of the electrode. A microelectrode is a fine-tipped (on the order of one micron across) conductor of electricity, typically a glass micropipette filled with a conductive fluid or a platinum or tungsten wire, that can be inserted into brain tissue (thus, a “penetrating electrode”), and is connected to an amplifier. Because it is conductive, it detects the changes in voltage that have been detailed in much of this chapter. Many of the properties of the neuron that have been described in this chapter were discovered via **intracellular recordings** obtained with in vitro methods, in which individual neurons, or more often slices of CNS tissue, are surgically removed from an animal and kept alive “in a dish” by bathing it in artificial CSF. Only by recording intracellularly can one measure, for example, the activity of a single postsynaptic channel, or the kinetics of voltage-gated ion channels. However, because cognitive neuroscience, by definition, deals with the neural processes that underlie behaviors (that we associate with cognition), the emphasis in this book will be on **extracellular recordings** obtained in vivo – that is, recordings made outside of the cell in the brain of a living animal.

From outside the cell, there are two types of signal that one can measure: action potentials and LFPs. Because action potentials are all-or-none events – there’s no such thing as a “big” or a “little” action potential – they are treated as binary events, and a standard way to quantify a neuron’s activity is by its firing rate. A neuron typically has a baseline firing rate, maybe 5 “spikes per second” (i.e., 5 Hz) for a generic pyramidal cell. The neuron’s involvement in a particular function, let’s say visual perception, can then be assessed by determining whether this firing rate changes systematically with a systematic change in the environment, say, flashing a light in an area that the animal is looking at. Depending on the neuron, and on the state of the brain, a neuron involved in the perception of this flash of light might exhibit either an increase or a decrease from its baseline firing rate. Depending on the cell type, an activated neuron’s firing rate can be anywhere from 15 Hz to upwards of 200 Hz.

To record the LFP, the electrode must be far enough away from any one cell body that its measurements are not dominated by the signal from that cell. Proper filtering of the signal removes the contribution from action potentials, leaving the local field potential – the aggregated fluctuation of dendritic membrane potentials from cells in the vicinity of the electrode (see *Figure 2.10* and *Web Clip 2.3*).

An alternative to recording electrical activity with penetrating electrodes is to use a grid of electrodes that is positioned on the surface of the cortex, a technique called **electrocorticography (ECoG)** (pronounced “ee-cog”) for short). Because ECoG electrodes lie on the surface of the cortex, they lack the spatial resolution that can be obtained with penetrating electrodes. Additionally, they can only measure field potentials. The unique knowledge gained with this method, however, is manifest in *Chapter 9* and *Chapter 18*.

influence the membrane voltage at the cell body. As we’ll see in *Chapter 4*, the biophysical properties of one common type of inhibitory interneuron make it tend to oscillate in the gamma range. A second factor that leads to the orderly oscillation of membrane

potentials is that many interneurons, in addition to sending axons to local pyramidal cells, are also “wired together” via axonal connections that they make with each other (i.e., they are reciprocally connected). As a result, the activity of these interneurons

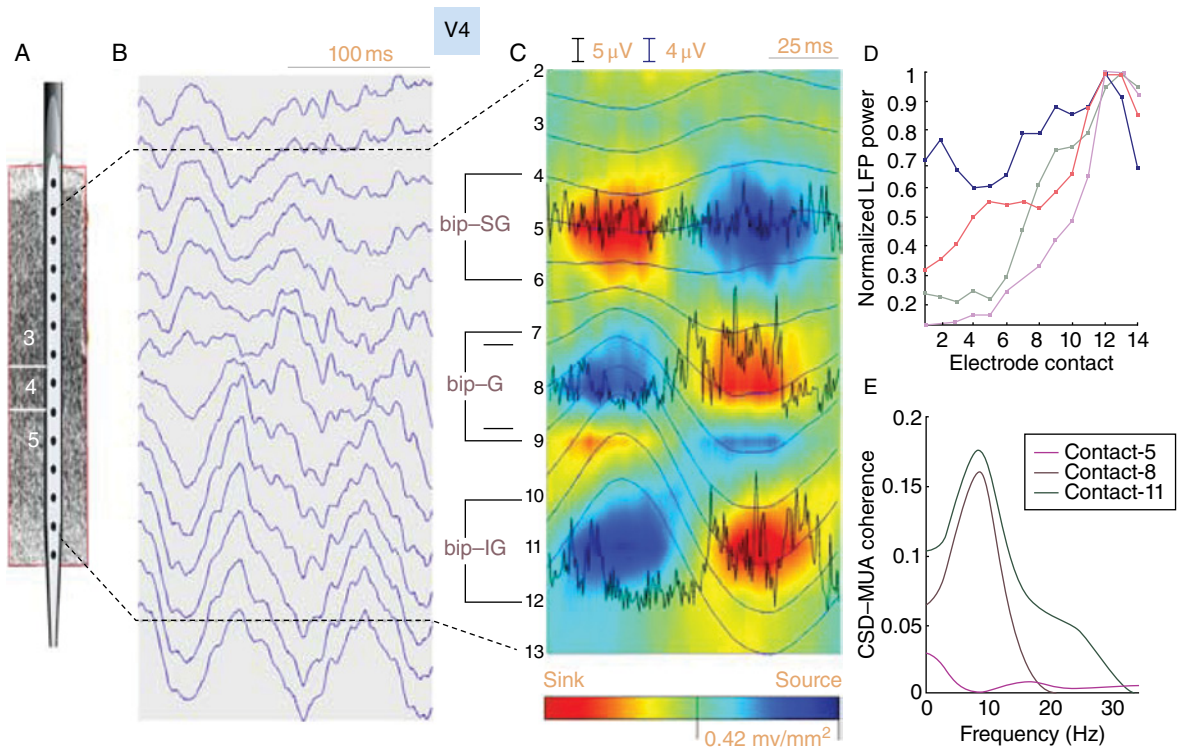


FIGURE 2.10 The local field potential (LFP). **A.** Schematic diagram showing an electrode with 14 equally spaced ($200\mu\text{m}$) contacts (“3,” “4,” and “5” refer to layers of cortex). **B.** 200ms of LFP recorded at each of these 14 contacts. **C.** A current source density map, computed from the pattern of LFPs in panel **B**. What this shows is the pattern, over time, of where in this cortical column current is flowing from intracellular space to extracellular space (a “source”), or from extracellular space to intracellular space (a “sink”). To compute the current source density map, bipolar recordings are needed, and “bip-SG” indicates the two contacts used to obtain these recordings in the supragranular layer, “bip-G” in the granular layer, and “bip-IG” in the infragranular layer. The jagged black traces show “multiunit activity” (MUA), a measure of voltage fluctuation due to the local firing of action potentials rather than to LFPs. The smooth blue lines are transformations of the raw LFP that we will disregard. These recordings were made in an area called “V4” (as referenced in *Figure 2.4*). (The experiment illustrated here is also discussed in *Web Clip 2.3*.) Source: Bollimunta, Chen, Schroeder, and Ding, 2008. Reproduced with permission of the authors.

tends to be synchronized. The consequence for pyramidal neurons is that they tend to experience a regular pattern of the delivery, then dispersion, of hyperpolarizing currents.

Synchronous oscillation

A consequence of receiving a synchronized delivery of hyperpolarizing GABA 40 times per second is that the membrane potential of a pyramidal cell won’t actually

remain at the steady -70mV that was implied at the beginning of this chapter. Rather, it will oscillate between, say, -90mV and -50mV . And because a neuron is much more likely to reach its firing threshold when its membrane potential is in the neighborhood of -50mV than that of -90mV , this results in a state in which the pyramidal neuron has, in effect, 40 “windows” per second during which it might be able to fire an action potential. Given this state of affairs, if a pyramidal neuron in area X wants to make a pyramidal

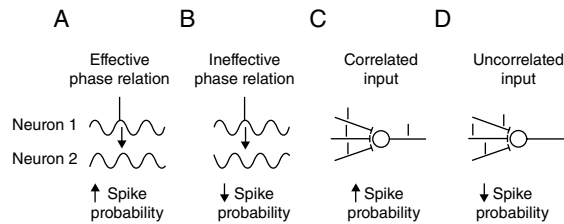


FIGURE 2.11 Illustration of how phase synchrony may govern the effectiveness of neural communication between a hypothetical “area 1” sending projections to a hypothetical “area 2.” Panels **A** and **B** illustrate synchrony between an upstream and a downstream neuron. In **A**, neuron 1 and neuron 2 are oscillating with a phase lag of near 0 (i.e., membrane voltage of neuron 1 reaches the maximum depolarization point of its cycle just a fraction of a second more quickly than does neuron 2). A consequence of this is that an action potential fired by neuron 1 (illustrated by “spike”) arrives at neuron 2 (illustrated by arrow) just as neuron 2 is at its maximally depolarized state, and thus the neuron 1 action potential has a relatively high probability of triggering an action potential in neuron 2. In panel **B**, neurons 1 and 2 are nearly 180 degrees out of phase (i.e., membrane voltage of neuron 1 reaches the maximal *depolarization* point of its cycle just a fraction of a second before neuron 2 reaches the maximal *hyperpolarization* point of its cycle). Thus, an action potential fired by neuron 1 arrives at neuron 2 just as neuron 2 is at its maximally hyperpolarized state, and thus the neuron 1 action potential has a relatively low probability of triggering an action potential in neuron 2. Panels **C** and **D** illustrate synchrony between several neurons in one region (“region 1”) that all project to a second region (“region 2”). In panel **C**, several upstream neurons are oscillating in phase with each other. Because the membrane potentials of neurons 1.A., 1.B., and 1.C. reach their maximally depolarized state at the same time, they are likely to fire action potentials at the same time. One consequence of this is that downstream neuron 2, to which 1.A., 1.B., and 1.C. all project, receives multiple depolarizing inputs all at the same time (illustrated by three “spikes” at same distance along three dendritic branches), and the summation of these inputs depolarizes neuron 2 to the point that it fires an action potential. In panel **D**, neurons 1.A., 1.B., and 1.C. are oscillating out of phase with each other, which means that their action potentials are not synchronized but, instead, randomly spread across time. The consequence for neuron 2 is that it receives depolarizing inputs that are staggered in time (illustrated by “spikes” at varying distances along three dendrites), and so there is little summation of inputs, thereby decreasing the likelihood that activity in region 1 will trigger an action potential in region 2. Source: Saalmann and Kastner, 2009. Reproduced with permission of Elsevier.

neuron in area Y fire an action potential, there are two things that it needs to do. First, it needs to time the delivery of its own action potential so that this action potential arrives during a depolarized phase of the membrane oscillation of neuron Y. Second, it needs many of its neighbors in area X to also send their action potentials at the same time, such that the area Y neuron receives many depolarizing impulses all at the same time. From this it is apparent that the synchronization of membrane oscillations in different populations of neurons is a critical factor in the smooth operation of the brain. Within a local group of cells, there’s a degree of synchrony between interneurons and pyramidal neurons. Across regions, oscillatory synchronization of **ensembles** of neurons is necessary if they are to effectively communicate. Indeed, in many

contemporary models of neuronal function, synchrony plays a critical role.

To illustrate how synchrony might be important for brain function, let’s walk through a hypothetical example. Let’s pretend that, while you’re walking through the woods, three things happen simultaneously: (a) the sun breaks through the clouds; (b) a bird in an overhead tree takes flight; and (c) a bear comes ambling along the path in your direction. Now, because individual elements in your brain don’t “know” that one of these events has more immediate relevance to you, the organism, than the other two, information about all three is propagated from the visual system to “downstream” brain systems that might implement one of three behaviors: (a) stopping to enjoy the warm caress of the sun; (b) reaching for the binoculars to watch the bird in flight; or (c)

RUNNING AWAY(!). According to synchrony models, one way that the nervous system has evolved to let choice (c) be the “winner” is that action potentials carrying information about the bear are timed to arrive during a depolarized phase in the oscillation of their downstream target system, whereas action potentials carrying information about the sun and about the bird arrive during a hyperpolarized phase of their downstream targets. Thus, the information about the bear is more potent, in a neural sense, and it thereby “wins” the “competition” to control your behavior. Such coordination and control of the LFP in different regions is referred to as **phase synchronization** (Figure 2.11). We shall see in subsequent chapters that some psychiatric disorders might be attributable to abnormal control of synchronization between neural systems.

It should be noted that whereas no neuroscientist questions the existence of oscillations in membrane potentials and in the LFP, there is vigorous debate about the extent to which these oscillations actually play a functional role in the workings of the brain. From one perspective, these oscillations could be nothing more than byproducts of the complex pattern of synaptic events experienced constantly by every neuron in the brain. This perspective might be paraphrased as *They happen, yes, but they may not actually serve any useful function*. On the other hand, with each passing month, it seems that there are new discoveries of specific neural systems (sometimes in the visual system (Chapter 3), sometimes in the systems that govern motor control (Chapter 7), sometimes in the hippocampus (Chapter 10)) whose functioning seems to depend on the precise synchronization of oscillations in different ensembles of neurons. Keeping this controversy in mind, the phenomenon of neuronal oscillations is one that we will return to at several points in this book.

COMPLICATED, AND COMPLEX

The human brain is a truly an amazing thing. It is made up of billions of neurons and trillions of synapses. What we have considered in this chapter only scratches the surface of all there is to know about the brain’s molecular and cellular makeup. Entire careers have been, and will continue to be, devoted to understanding these elemental components of the brain. With the facts that we have considered, and the questions we have posed, however, we have enough foundational knowledge that we can now get on to what this book is about: cognition. While all these facts about how the brain is put together and how it works are still fresh in the mind, however, it is worthwhile considering a distinction made very eloquently by György Buzsáki in his 2006 book *Rhythms of the Brain*. The nuts-and-bolts facts of how cells in the brain work are, indeed, *complicated*. How this all works together to give rise to cognition, however, is not only complicated, it’s *complex*. “The term ‘complex’ does not simply mean complicated but implies a nonlinear relationship between constituent components,” and refers to properties often termed “spontaneous, endogenous, ... self-organized, self-generated, self-assembled, and emergent” (p. 11). Stated another way, what we shall see again and again in our exploration of cognitive neuroscience is that “the whole is greater than the sum of its parts” (see *Thought Box 1.2*). This chapter has introduced many of the critical “parts” – parts of the brain, neurons, and neurotransmitters – as well as some of their properties – ion channels, oscillating membrane potentials, action potentials, and oscillating LFPs. Now we can start to explore how these combine and interact to produce the wonder that is cognition. And we’ll start with visual perception.

END-OF-CHAPTER QUESTIONS

1. In what ways do the anatomical organization of the adult brain follow from its developmental origin as the neural tube?
2. What factors explain the “crumpled up” shape of the brain? What evolutionary pressures may have produced this shape?
3. What is a distinctive property of each layer of neocortex? Why might there be some advantages to segregating these properties/functions by layer?
4. What are the principal disequilibria of the neuron that were emphasized in this chapter, and what role do they play in the generation of the action potential?

5. How are the disequilibria raised in question 4 maintained by the neuron? In what way does your answer to this question relate to physiological signals that are detected with fMRI and some types of functional tissue imaging?
6. It can be said that neurons communicate via signals that, at one level of analysis can be considered digital, yet from another level of analysis can be considered analog. What phenomena that were discussed in this chapter are an example of each?
7. What explains the oscillation of voltage across the cell membrane? What explains the oscillation of voltage in the LFP? How are these two types of oscillation measured?
8. Although some scientists have argued that neuronal oscillations are primarily byproducts of the workings of neurons and networks of neurons (i.e., that they may be “epiphenomenal”), there is a growing consensus that they may play an important role in governing the effectiveness of communication between one brain area and another. Describe the concept of phase synchronization, and the role it may play in neuronal communication.

REFERENCES

- Bollimunta, Anil, Yonghong Chen, Charles E. Schroeder, and Mingzhou Ding. 2008. “Neuronal Mechanisms of Cortical Alpha Oscillations in Awake-Behaving Macaques.” *Journal of Neuroscience* 28 (40): 9976–9988. doi: 10.1523/JNEUROSCI.2699-08.2008.
- Buszáki, György. 2006. *Rhythms of the Brain*. Oxford: Oxford University Press.
- Dumitriu, Dani, Alfredo Rodriguez, and John H. Morrison. 2011. “High-Throughput, Detailed, Cell-Specific Neuroanatomy of Dendritic Spines Using Microinjection and Confocal Microscopy.” *Nature Protocols* 6 (9): 1391–1411. doi: 10.1038/nprot.2011.389.
- Hubel, David H., and Torsten Wiesel. 1977. “Functional Architecture of Macaque Monkey Visual Cortex.” *Proceedings of the Royal Society of London. Series B, Biological Sciences* 1989 (1130): 1–59.
- Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell, eds. 2000. *Principles of Neural Science*. New York: McGraw-Hill.
- Petrof, Iraklis, Angela N. Viaene, and S. Murray Sherman. 2012. “Two Populations of Corticothalamic and Interareal Corticocortical Cells in the Subgranular Layers of the Mouse Primary Sensory Cortices.” *Journal of Comparative Neurology: Research in Systems Neuroscience* 520(8): 1678–1686. doi: 10.1002/cne.23006.
- Saalmann, Yuri B., and Sabine Kastner. 2009. “Gain Control in the Visual Thalamus during Perception and Cognition.” *Current Opinion in Neurobiology* 19 (4): 408–414. doi: 10.1016/j.conb.2009.05.007.
- Saalmann, Yuri B. and Sabine Kastner. 2011. “Cognitive and Perceptual Functions of the Visual Thalamus.” *Neuron* 71 (2): 209–223. doi: 10.1016/j.neuron.2011.06.027.
- Wandell, Brian A., and Jonathan Winawer. 2011. “Imaging Retinotopic Maps in the Human Brain.” *Vision Research* 51(7): 718–737. doi: 10.1016/j.visres.2010.08.004.
- Warwick, Roger, and Peter L. Williams. 1973. *Gray's Anatomy*. 35th British ed. Philadelphia: W. B. Saunders.

OTHER SOURCES USED

- Cooper, Jack R., Floyd E. Bloom, and Robert H. Roth. 1991. *The Biochemical Basis of Neuropharmacology*, 6th ed. New York: Oxford University Press.
- Diamond, Marian C., Arnold B. Scheibel, and Lawrence M. Elson. 1985. *The Human Brain Coloring Book*. New York: Harper Perennial.
- Hubel, David H. 1988. *Eye, Brain, and Vision*. Scientific American Library, no. 22. New York: W. H. Freeman.
- Oztas, Emin. 2003. “Neuronal Tracing.” *Neuroanatomy* 2: 2–5. http://www.neuroanatomy.org/2003/002_005.pdf.

FURTHER READING

Buszák, György. 2006. *Rhythms of the Brain*. Oxford: Oxford University Press.

If there's only one book that this textbook inspires you to read, this might be it. Provides a rigorous neurophysiological and anatomical basis for much of the dynamical systems-influenced content of this book, including, for example, the final section of this chapter.

Diamond, Marian C., Arnold B. Scheibel, and Lawrence M. Elson. 1985. *The Human Brain Coloring Book*. New York: Harper Perennial.

In science, it's often said that "the best way to ensure that you understand a topic is to teach a class about it." The same might be said about neuroanatomy and drawing.

Morrison, John H., and Mark G. Baxter. 2012. "The Ageing Cortical Synapse: Hallmarks and Implications of Cognitive Decline." *Nature Reviews Neuroscience* 13 (4): 240–250. doi: 10.1038/nrn3200.

In Chapter 14 we'll learn that the prefrontal cortex is one of the first regions of the brain to show age-related decline in healthy, middle-aged individuals. This paper shows that, in the monkey, this decline is largely due to a dramatic loss of dendritic spines in layer III, the cortical layer whose integrity is most closely tied to age-related cognitive impairment.

Oztas, Emin. 2003. "Neuronal Tracing." *Neuroanatomy* 2: 2–5. http://www.neuroanatomy.org/2003/002_005.pdf.

A very concise overview of some of the physiology, and biochemistry, underlying anatomical tracing techniques.

Wandell, Brian A., and Jonathan Winawer. 2011. "Imaging Retinotopic Maps in the Human Brain." *Vision Research* 51(7): 718–737. doi: 10.1016/j.visres.2010.08.004.

Nice treatment of the use of MRI for tractography. Additionally, some of the content of this paper, particularly in the section on "Functional specialization and maps," will be relevant in subsequent chapters, particularly Chapter 9.

SECTION II

SENSATION, PERCEPTION, ATTENTION, AND ACTION

INTRODUCTION TO SECTION II SENSATION, PERCEPTION, ATTENTION, AND ACTION

Near the end of *Chapter 1* we discussed the fact that brains confer the ability to detect changes in the environment, and to perform actions based on this information. The chapters in this section will delve into the wondrous amount of intricate detail and organization that underlies this simple premise. As a preamble, let's briefly consider the four words in the title of this section, the concepts that they encompass, and how these classes of function relate to one another.

If we think of the functions of the brain as lying along a detection-to-action continuum, **sensation** is the term given to the neural processes that correspond most closely to the concept of detection. Under this rubric falls the translation of information from the environment into a neural signal that includes a representation of the physical properties of the environmental information (e.g., brightness for vision, loudness and pitch for audition, force and texture for touch, and concentration for olfaction). **Perception** refers to the identification of features of what is being

sensed, and, often, to its recognition (e.g., *Am I looking at a house cat or a raccoon?*, *Is the voice on the telephone that of my daughter or of my wife?*, *Does the blade of this ice skate feel sufficiently sharp as I run my finger along it?*, *Does this milk smell as though it has gone sour, or can I drink it?*).

Attention can refer to many things, but in the present context it's how we mentally prioritize the perception (and related actions) of just one or a few of the multitudes of stimuli that typically clutter our environment. In *Chapter 2*, oscillatory synchrony was introduced as a mechanism for focusing one's attention on an approaching bear rather than on the sunlight or on a bird. **Action** can refer to isolated acts of motor control (e.g., reaching for the computer mouse while surfing the Web, or turning one's head to see who it is who is walking into the room), but also to goals or plans that can be abstracted from isolated movements, such as *making dinner* or *finishing my homework*.

The definition of sensation has raised a concept that will be of central importance to all of the

chapters of this section, and that's the idea of translating information into signals that the nervous system can understand and manipulate. This is fundamental, because when you see a cat, for example, it's not the case that the photons reflecting off this animal are traveling into your brain to be projected against a "neural screen." Nor is it the case that, when you smell bread baking in the oven, molecules of heated dough are entering your brain through your nose. Rather, the nervous system represents information about the seen cat and about the smelled bread with what we call a **neural code**. A code is a system for representing information in a format that's different from the physical properties of what's being represented. The English language is a code by which, for example, this author can convey to you, the reader, the idea of a cat without having to plop a furry, domesticated feline into your lap. Similarly, the nervous system uses a code made up of action potentials and neurotransmitters to convey information about patterns of light or the strength and quality of a smell. It uses this same code to trigger the click of your finger on the computer mouse, and the speed and manner (quickly and with alarm, or slowly and with insouciance) with which you turn your head toward the door. There is much about the neural code that we do not yet understand, and discovering the rules that govern it is an important goal for cognitive neuroscience. (By analogy, without knowledge of the rules governing the English language, one wouldn't be able to understand the difference between the sentences "Dog bites man" and "Man bites dog.") At the front end of "translating information into a neural code" is the translating part, that is, the **transduction** of physical energy into the neural code. For vision, we'll consider how the retina transduces light of different wavelengths; for audition, how the cochlea transduces fluctuations in air pressure; for somatosensation, how receptors in the skin transduce different mechanical forces, chemical agents, and temperatures that the body encounters.

It's a matter of organizational convenience that this section presents the topics of *sensation*, *perception*,

attention, and *action* in the order that it does. However, an important fact to keep in mind as one reads through these chapters is that it would be overly simplistic, to the point of being inaccurate, to think of neural information processing as proceeding in a strictly feedforward series of stages from sensation to perception to attentional selection, to action. We saw in *Chapter 2* that the cortex is wired up for the feedback as well as for the feedforward conveyance of signals, and in *Chapter 3* we'll begin to see some of the functional consequences of this. A related point is that it would be equally inaccurate to think of the brain as "sitting passively" until a stimulus comes along to "activate it," to kick-start it into action. Instead, we'll see that a better model might be to think of a brain that is constantly "churning" along, actively representing the world, and that the sensation of a newly presented stimulus corresponds to the "nudging" of this already-active system into a different state of activity. That is, the brain is a dynamic system, and the waking brain is never truly "at rest." I raise these points now in part because, even though I believe them myself, I often find myself slipping into a more straightforward, intuitively appealing way of thinking about brain function. By this intuitive account, a newly occurring stimulus evokes activity in a previously quiescent brain, and this event triggers the sensation-through-action cascade that has been previewed in this preamble. I also raise these points to be explicit about the fact that this "feedforward activation" perspective, by virtue of its intuitive appeal, offers a more straightforward way of introducing many aspects of brain function. This is evident in much of the terminology that we use to describe the brain and its functions, in figures that appear in many textbooks, and, indeed, in the order of chapters in textbooks such as this one. What we shall see in this section, however, is that a feedforward-activation view of brain function can limit one's ability to appreciate all its nuances. To give just one example, what one perceives can often be influenced, if not determined, by one's action plan.

But that's enough "pre"; let's dive in!

CHAPTER 3

SENSATION AND PERCEPTION OF VISUAL SIGNALS

KEY THEMES

- The brain's representation of the visual world is dictated by the optics of the eye and, in particular, where light from the visual scene falls on the retina.
- The topography of the retina is reflected in the retinotopic organization of primary visual cortex.
- The disproportionate retinal representation of central (i.e., foveal) vision results in its cortical magnification in terms of the surface area of V1.
- A central principle for understanding the response properties of retinal photoreceptors, retinal ganglion cells, neurons of the lateral geniculate nucleus (LGN) of the thalamus, and neurons of primary visual cortex (area V1), is that of the receptive field.
- Individual neurons of V1 assemble input from multiple LGN neurons to construct specific features to which they respond preferentially, and thereby act as feature detectors.
- At least five classes of features emerge at the level of V1: orientation; direction of movement; higher acuity; binocular disparity; end-stopping; and the representation of color.
- Functional tissue imaging, which permits visualization of a large surface of cortex, has revealed that columns representing smooth transitions through the full 180° of orientation are organized in pinwheel-like configurations.
- Visual information processing does not consist of simply a serial feeding forward of signals from retina to LGN to V1, because V1 sends heavy feedback to the LGN, as do higher-level visual areas to V1.
- Functions of feedback include biasing the LGN to emphasize the orientation vs. the direction-of-movement of visual stimuli, and determining how effective signals from LGN will be at driving cortical responses.
- At a broader spatial scale, the processing of incoming visual information is influenced by the magnitude of ongoing oscillations in the alpha frequency band (~9–13 Hz) of the EEG.
- A principle underlying the overall functioning of the visual system may be one of predictive coding – the construction and continual updating of an internal model of what is likely to happen in the future.

CONTENTS

KEY THEMES

THE DOMINANT SENSE IN PRIMATES

ORGANIZATION OF THE VISUAL SYSTEM

The visual field

The retinotopic organization of primary visual cortex

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – BOTTOM-UP FEATURE DETECTION

The V1 neuron as feature detector

Columns, hypercolumns, and pinwheels

Behind the pinwheel

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – INTERACTIVITY

Feedforward and feedback projections of V1

*Corticothalamic loops, center-surround
inhibition, and signal-to-noise ratio*

*Setting the state of thalamic neurons:
tonic mode vs. burst mode*

Circularity? It can depend on your perspective

The relation between visual processing and the
brain's physiological state

WHERE DOES SENSATION END? WHERE DOES PERCEPTION BEGIN?

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

THE DOMINANT SENSE IN PRIMATES

Of all our senses, one could make a strong case that vision is singularly important for successfully navigating through and interacting with our environment. (As a thought experiment, imagine being dropped off in an unfamiliar city, with no one to guide you, while either (a) blindfolded, (b) “earplugged,” (c) “nose-plugged,” etc.) It is also probably the case that vision is the sense that has been most studied from each of the three traditions that have informed contemporary cognitive neuroscience most heavily: systems neuroscience; neuropsychology; and experimental psychology. As discussed in *Chapter 1*, visual perception is a function characterized by close linkage between the environment and the brain. Indeed, a useful first step in considering the cognitive neuroscience of vision is to learn the conventions of how scientists describe this interface.

ORGANIZATION OF THE VISUAL SYSTEM

The visual field

The **visual field** refers to the spatial extent of everything that you can see right now. While you’re reading this page, it’s likely that your visual field is largely filled by this page and the one opposite, with blurry images of what’s behind the book surrounding the more in-focus percept of the book itself. Note that when you pay attention to the sharpness-of-focus of your vision for words that are one or two lines above or below where you’re currently reading, it’s clear that your acuity falls off quite quickly in all directions. The center of gaze, where your acuity is sharpest, is referred to as foveal vision, because it corresponds to the part of your retina called the **fovea** (see *Thought Box 3.1*). Now move your center of gaze over to the crease that’s formed between the left and right page of the book. Assuming that you’re holding the book upright, this crease is passing through your fovea and defining the boundary between the left side and the right side of your visual field. The region to the left is referred to as the left **visual hemifield**, and the region

to the right is the right visual hemifield. Note that this remains true regardless of the position of your eyes in your head. To demonstrate, start by holding your book in front of you and fixating the central crease (to fixate is to hold one’s center of gaze on a selected location, often on an object that is stationary at that location). Now, while holding the book still and holding your head still, move your eyes as far to the left as you can, and hold them there, your center of gaze falling on an object to the left of the book. The book is still in your field of view, but it now occupies the right visual hemifield. Thus, your visual field moves around with your eyes. Conversely, if you keep your center-of-gaze fixated on one point while slowly moving your head back and forth, or up and down, your visual field remains stationary.

Now move your eyes back to the vertical crease dividing the left vs. right page of the book. The imaginary line that lies on top of the crease and extends to the uppermost and lowermost reaches of the visual field is called the **vertical meridian** of the visual field. In the same way that one can vertically divide the visual field into left vs. right hemifields, one can also divide it into upper vs. lower hemifields, this divide defined by the **horizontal meridian**. Together, the two meridians divide the visual field into quadrants: upper right, lower right, lower left, and upper left.

Before we start to consider how these components of the visual field are mapped onto the brain, let’s try one more simple exercise. Close your left eye. Note that, with this eye closed, although your visual field is somewhat narrower than it is with both eyes open, you can still see much of the left visual hemifield. That is, *the two hemifields do not correspond to each of the two eyes*. Rather, each eye represents most of the visual field, with considerable overlap in the center, and only the most eccentric (i.e., farthest-from-the-center) portions of each hemifield are only represented by one eye. The reason that one eye can represent most of the visual field follows from its structure. If one thinks of the eye as a transparent globe, the retina, which is the portion of the eye that transduces light energy into neural signals, is attached to the inside surface of the half of the globe that’s inside the eye socket. That is, it is concave. Thus, as illustrated in

THOUGHT BOX

3.1 How does the retina work, and why is it organized as it is?

How does it work?

Phototransduction, the process of converting light into neural signals, is accomplished by a class of cells called photoreceptors. In a sentence, these cells contain a pigment called rhodopsin that, when impacted by light, undergoes a conformational change that triggers an intracellular molecular cascade resulting in the closing of membrane-spanning ion channels. Thus, a type of physical energy (electromagnetic radiation with wavelengths between 400 and 700 nm (i.e., “visible light”)) is converted into an electrical signal, in this instance, a change in the membrane potential of the photoreceptor cell. This change in membrane potential produces currents within the complex cellular circuitry of the retina, the end result of which is that action potentials are generated by **retinal ganglion cells**, the output neurons of the retina whose axons project to the LGN.

There are two classes of photoreceptors, cones and rods. Cones are sensitive to color, with the rhodopsin in “blue,” “green” and “red” cones being preferentially sensitive to light at short, medium, and long wavelengths, respectively. Rods, on the other hand, are not sensitive to wavelength (i.e., are “colorblind”), and contribute primarily to vision at extremely low levels of illumination, such as at night.

Organization?

Although containing an array of detector elements, not unlike the grid of photodiodes in a digital camera (see *Tip 3.1*), the packing density of photoreceptors in the retina is not uniform. Rather, at the fovea, one finds the highest density of cones, which drops off precipitously as one moves in any direction toward the periphery. For this reason, the fovea provides the highest acuity, the highest spatial resolution, of any part of the retina. This explains why, for example, we move our eyes across the page as we read: Only the highest-acuity portion of the retina can deliver the detail that is needed to resolve small type. The same is true for recognizing faces, for appreciating a painting, for reading road signs, etc. As one moves away from the fovea toward the periphery, three properties begin to change. One is the progressive decrease in the density of cones. A second is that the surface area that each cone occupies on the retina increases. A third is that an increasing number of rods surround each cone. It is a consequence of the fact that there are fewer cones devoted to each square unit of surface in the periphery of the visual field that our peripheral vision is of much lower resolution than is foveal vision.

A reasonable supposition as to why the retina is organized this way comes from considering the incredible amount of processing power required to analyze information from the fovea (as is detailed in the main text, beginning with the consideration of cortical magnification). If the packing density of the entire surface of the retinae were the same as that at the fovea, it's been estimated that one's optic nerves would need to be as thick as one's arms (to accommodate so many more axons), and one's primary visual cortex would need be the size of a large room (to accommodate all the neurons needed to process so much high-resolution visual information).

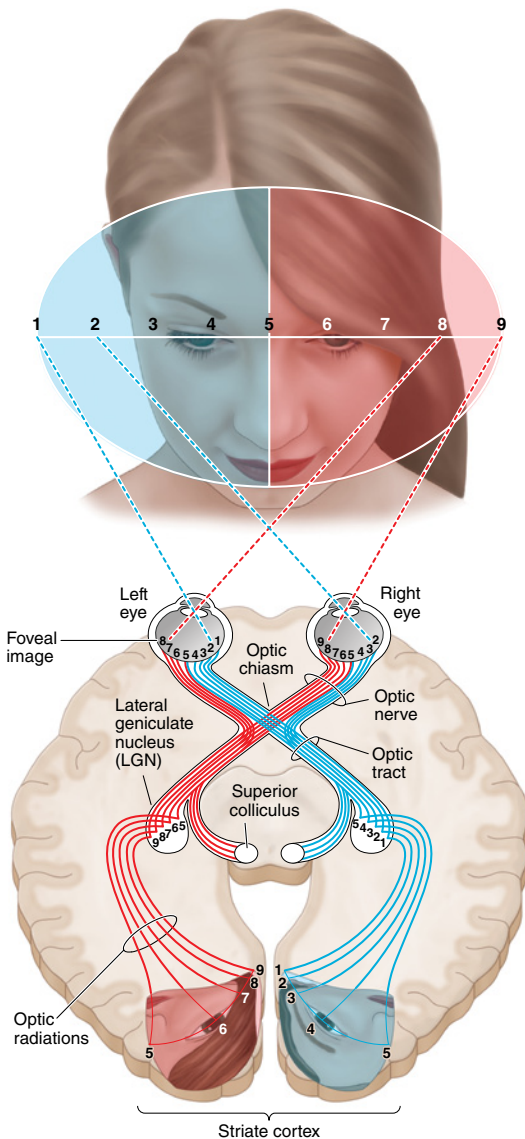


FIGURE 3.1 Illustration of the visual field, emphasizing the right (shaded red) and left (shaded blue) hemifields, and how information from these hemifields is conveyed to primary visual cortex (labeled as “striate cortex,” for reasons explained in the text). Source: Wolfe, Kleunder, Levi, Bartoshuk, Herz, Klatzky, Lederman, and Merfeld, 2012. Reproduced with permission of Sinauer Associates, Inc.

Figure 3.1, light that is emanating from the left side of the visual field strikes the right side of the retina, and light that is emanating from the right side of the visual

Tip 3.1

The word “pixel” is an abbreviation for “picture element.” The resolution of digital cameras is typically quantified in terms of “megapixels” (1 megapixel = 1 million pixels). The human retina has roughly 6 million cone cells (perhaps better referred to as “sensors,” because they are sensory elements, rather than display (or “picture”) elements). How does the resolution of your retina compare to that of your digital camera?

field strikes the left side of the retina. For there to be a “one eye = one visual hemifield” arrangement, one of two things would have to be different about our eyes. Either (1) the retinæ would have to be flat, instead of concave; or (2) the eyes would have to be farther apart, halfway between their present location and the temples. Animals that are “wall-eyed,” like most fish and birds, have two nonoverlapping visual fields.

The retinotopic organization of primary visual cortex

We concluded *Chapter 1* with three branches of science – systems neuroscience; neuropsychology; experimental psychology – poised to make important discoveries about brain structure and function at the dawn of the twentieth century. With regard to vision, once occipital cortex was defined as its necessary substrate, the next task was to work out its organization. The fact that there is systematicity to the organization of visual cortex was soon established by neuropsychologists and neurologists whose studies of soldiers with penetrating head injuries indicated that damage to specific areas of the occipital lobe invariably produced partial blindness in the same regions of the visual field (*Tip 3.2*). Thus, for example, damage to dorsal regions of the right occipital lobe produced blindness in the lower-left quadrant of the visual field. A large lesion would produce a quadrantanopia – blindness in the entirety of the

Tip 3.2

Advances in our understanding of the visual system were spurred by introduction of high-velocity bullets during the Russo-Japanese war, which resulted in numerous casualties with discrete brain lesions and clear entry and exit wounds in the skull that facilitated localization of the damage. Thus, Japanese ophthalmologist Tatsuji Inouye (1880–1976) made pioneering discoveries about retinotopic organization and cortical magnification in V1. World War I spurred similar discoveries in Western Europe (see *Figure 3.5*).

Tip 3.3

Primary visual cortex is also known as area V1, the anatomically adjacent area receiving feedforward projections from V1 as area V2, and so on. The same convention is true for the auditory system (A1, A2, ...) and the somatosensory system (S1, S2, ...). Note that the serial ordering of brain regions implied by this numbering scheme doesn't extend beyond "2," in that V1 also has direct projections to area V5, and V4 and V5 are considered to occupy analogous positions along two separate pathways.

quadrant – or a hemianopia – blindness in one half of the visual field – whereas a smaller lesion might produce a **scotoma** within that quadrant. An implication of this finding was the realization that everything in the primary visual cortex is “flipped” with respect to the visual field. Things that appear to the left of the vertical meridian are initially processed in the right hemisphere of the brain, and things that appear above the horizontal meridian are initially processed in ventral portions of the occipital lobe. To understand why, we must turn to the anatomy of the visual system.

Both “flips” can be understood as resulting from the optics of the eye. For now, let's keep the left eye closed, and consider just the right. The concave shape of the retina on the back of the eyeball dictates that light reflecting off stimuli appearing below the point of fixation will be projected onto the upper retina, and that stimuli appearing to the right of fixation will be projected onto the left side of the retina. Because we're only considering the right eye for the moment, this left half of the retina is referred to as the nasal retina, because it is closest to the nose. (The half of the retina that is closest to the temple is the temporal retina.) And so, already in the eye, what's “up” in the world is “down” in the retina, and what's “left” in the world is “right” in the retina. This flipped state of affairs is preserved as axons from the retina project posteriorly to the thalamus. Along the way, the fibers that correspond to the nasal retina cross over to the left hemisphere, so that half of the axons from the right eye – those from the temporal retina – project to the right lateral geniculate nucleus (LGN) of the

thalamus, and the half from the nasal retina project to the left LGN (*Figure 3.1*). Another important fact about these axons is that their trajectory preserves the layout of the retina, such that axons from adjacent retinal ganglion cells synapse onto adjacent neurons of the LGN. Similarly, adjacent neurons of the LGN project to adjacent neurons in primary visual cortex (a.k.a. V1; *Tip 3.3*). In this way, the spatial organization of V1 reflects the spatial organization of the retina (*Figure 3.1*). If you're looking at the trunk of a tree, green light from the leaves is reflecting onto the lower half of the retina, and the dirt where the trunk meets the ground is reflecting onto the upper half of the retina. When V1 receives this information via the LGN, ventral portions of this region begin processing information about the color, texture, etc. of the leaves, whereas dorsal portions of this region begin processing information about the dirt. Thus, the functional topography of V1 mirrors the functional topography of the retina. This is referred to as **retinotopic organization**, or **retinotopy**.

Inspection of *Figure 3.1* and *Figure 2.6* identifies the fiber pathways that convey information from the eye to the cortex. (Note that the retina is considered a part of the brain, and so I'm being careful not to say “from eye to brain.”) Although all axons project uninterruptedly from retina to LGN, their bundles from eye to optic chiasm are referred to as the *optic nerve*, whereas downstream of this decussation (“crossing over”) it's the *optic tract*. Thus, the optic nerves carry eye-specific information, whereas the optic tracts carry visual field-specific information (see *Tip 3.4*).

Tip 3.4

The loss of V1 in its entirety in one hemisphere produces hemianopia: blindness in the corresponding visual field. To get a feel for what this means, start by again closing your left eye. At this point you're not yet mimicking a hemianopia. Rather, you're mimicking being blind in one eye. To simulate a hemianopia, you now need to hold one open hand up to your nose, just as you would to block a two-finger eye-poke from one of the Three Stooges. For a better effect, use something even longer, like both hands, or this book. What you are doing now is depriving your visual system of information from the left visual hemifield.

The LGN itself gets its name from the fact that an early anatomist thought it looked like a bent knee, and geniculate derives from Latin for “knee-like” (see *Tip 3.5*). The fibers of the optic tract terminate in one of several layers of the LGN, according to a principle that we'll return to further along in this chapter. In the human, the projection from LGN to V1 runs along the dramatically curving and fanned-out optic radiations, a striking consequence of the mechanical twisting and stretching of this fiber tract brought about by the rapid growth of the brain during in utero development (*Figure 2.6*). Note that these radiations terminate along the medial surface of the occipital lobe. More specifically, the majority of them terminate inside a sulcus that runs parallel to the rostro-caudal axis of the brain, starting near the occipital pole and extending for roughly the length of the occipital lobe. This is the calcarine fissure, named after the shape that its indentation makes in the adjacent posterior horn of the lateral ventricle (see *Tip 3.6*).

The axial view depicted in the MR image in *Figure 2.6* is not dissimilar to the one that Italian medical student Franciso Gennari inspected, and later described in a 1782 book, noting a thick white stripe in cross-sectional views of the cortex lining the calcarine fissure. This “stripe of Gennari,” as it has come to be known, was later shown to correspond to myelinated fibers coursing through layer IV of this cortex. Its primary significance for this book is that the term

Tip 3.5

We'll see over and over that the study of the brain provides a mini crash course in Latin: *amygdala* is Latin for almond, *hippocampus* is Latin for seahorse, *substantia nigra* Latin for “black substance” ...

Tip 3.6

Although all sulci are indentations, most just indent into underlying white matter. The calcarine fissure, however, indents into the floor of the lateral ventricle. Viewed “from above,” this resembles the spur of a rooster's leg lying on the floor of the ventricle (thus, in Latin, *calcar avis*). Chapter 4 of Gross's *Brain, Vision, Memory* (1998) gives a fascinating account of how this insignificant feature of the brain played a central role in nineteenth-century debates over evolutionary theory and Man's place in nature.

striate cortex (i.e., “striped cortex”) has come to be one of the most common terms according to which this region is known. Relatedly, the visual areas “outside” of V1, but to which it projects, are known collectively as **extrastriate cortex**.

Another consequence of the fact that the organization of V1 is dictated by the organization of the retina is that roughly 80% of V1 is devoted to processing information from the central 10% of the retina, a simple consequence of the inhomogeneity of cone density on the retinal surface (*Thought Box 3.1*). What this means for V1 is that there is a much larger portion of cortical area devoted to processing information from the fovea than from the peripheral retina, thereby leading to the distortion of the retinal image that is “projected” onto the cortex (*Tip 3.7*). That is, columns of neurons are spaced the same distance apart at the rostral-most extent of the calcarine fissure as they are at the occipital pole. Thus, if there is to be a one-to-one correspondence between photoreceptors and V1 neurons (which strictly speaking, there isn't, but the approximation is fine for our purposes), it's clear why

Tip 3.7

A technically more accurate way to write the phrase “the retinal image that is ‘projected’ onto the cortex” would be “the retinal information that is represented in the cortex to which the retina (via the LGN) projects.” Cognitive neuroscientists often take expository shortcuts, such as this one, and so it is important when reading (and writing) to know when a phrase is intended metaphorically (as it is here), and when it is meant to be taken literally (not here, because there is no actual *image* being *projected* onto a cortical surface).

the cortical representation of the fovea, with its many more columns of neurons, covers a large area relative to the cortical representation of the periphery, with its relatively fewer columns of neurons. Such disproportionate representation of high-acuity portions of a sensory system is referred to as **cortical magnification**. This property is evident in the distorted retinotopy illustrated in *Figure 3.5*, and will be seen again when we consider other systems, prominently so in the somatosensory (*Chapter 4*) and motor (*Chapter 7*) systems.

An additional implication of retinotopy for V1 neurons is that they will be sensitive to stimulation within a larger or smaller portion of the visual field, depending on what portion of the retina a neuron represents. This fundamental property of a sensory neuron is referred to as its **receptive field** (*Thought Box 3.2*). In V1, receptive fields are smallest at regions receiving input from the fovea, and largest at regions receiving input from the periphery, properties that follow from the spatial resolution of foveal vs. the peripheral regions of the retina (*Thought Box 3.1*). Because a V1 neuron’s receptive field properties are determined by the region of the retina that projects to it, the most direct way to quantify the dimensions of its receptive field would be to do so in terms of the surface area of retina from which it is receiving input. Thus, as an approximation, one might estimate that a V1 neuron representing the fovea has a receptive field corresponding to a 5×10^{-5} mm² surface area of the retina, whereas a V1 neuron representing the periphery

might have a receptive field corresponding to a 5×10^{-4} mm² surface area of the retina. Although these values reflect the causal factors determining V1 receptive field properties, they are neither intuitively graspable, nor very meaningful with regard to the function of primary visual cortex. Instead, it is much more useful to think of receptive field size in terms of **degrees of visual angle**.

To get a feel for the concept of visual angle, set down your book for a moment and fixate a point on the wall across the room. Next, hold your arms out straight to each side, as though you were about to start flapping them like wings. Now slowly move your arms toward each other, as though you were going to clap your hands with locked elbows (in the way that people do to mimic a trained seal) and stop moving them when each hand is just visible in the right and left periphery of your field of view. A person with normal vision in both eyes will find that her arms don’t get far before she has to stop. As you hold this position with arms slightly bent at each shoulder, you can think of your arms as defining a very wide angle, let’s estimate it to be 170°. Next, swing your right arm into your field of view until it is pointing at the same spot that you are fixating. The imaginary line passing through your hand and stretching from floor to ceiling defines the vertical meridian, and now your right and left arms are defining an angle that is one half of 170°, or 85°. And so, by these rough calculations, the visual field subtends 170° of visual angle along the horizontal meridian. Finally, let’s make one more observation before you relax your right arm. With your elbow straight, hold up your thumb (as if indicating “1” on Continental Europe, or signaling “I am cool” in the United States). As you stare at it, the width of the joint in the middle of the thumb subtends approximately 2° of visual angle. Having established this more functional metric of size in the visual system, we can return to receptive fields in V1. In regions of V1 representing the fovea, receptive fields can be as small as $0.5^\circ \times 0.5^\circ$ of visual angle, or smaller, whereas as in regions of V1 representing the periphery they can be an order of magnitude or more larger.

At this point, we’ve considered some principles of the early stages of visual sensation: how light projects onto the retina and how it is transduced into a neural code; how visual signals are carried to the cortex; and

THOUGHT BOX

3.2 The receptive field

Two adjacent photoreceptors in the retina are likely to receive information from adjacent portions of a visual scene. For example, when you're reading the word "adjacent," there is a viewing distance at which photons reflecting off the dot over the *j* are impacting one photoreceptor and the photons reflecting off the lavender-colored space between the dot and the letter are impacting a photoreceptor just above it. These two photoreceptors have different receptive fields, in that they *receive* information from different portions of the visual *field*. Now, let's make the dot on the letter *j* blink on and off. If your eye were perfectly still, the photoreceptor whose receptive field is restricted to the tiny lavender space under the dot wouldn't detect anything about the blinking dot. That's because this dot is not in this photoreceptor's receptive field. The principle of the receptive field holds for retinal ganglion cells, and for neurons of the LGN and V1 (indeed, for many other areas of the brain).

One feature of V1 receptive fields is that they are larger than those of foveal photoreceptors. For example, a V1 neuron with a receptive field centered on the tiny lavender space below the dot of the *j* might also encompass the area corresponding to the dot. The reason for this is that each V1 neuron receives input from several retinal photoreceptors (*Figure 3.3*). One consequence of such "pooling" across the inputs from several retinal photoreceptors is that a V1 neuron necessarily has a lower spatial resolution than do any of the individual photoreceptors that feed into it. A simplified way to think about this is as follows: If five photoreceptors feed into a single V1 neuron, and that V1 neuron fires each time any of these five photoreceptors signals that something has changed in its receptive field, anyone "listening" to that neuron can know when something has happened within the area covered by those five photoreceptors, but can't know precisely where within that area of five photoreceptor receptive fields the change has occurred.

There is, however, an important tradeoff for the loss of spatial resolution that occurs between the retina and V1, and that is a gain in information about *what* has changed in the visual world. In effect, by sampling across the input of multiple photoreceptors, neurons in V1 can "infer" information that was not contained by any one photoreceptor, alone. This process is already evident at the level of the retinal ganglion cell. Although it may fire robustly when a tiny spot of light is projected into the center of its receptive field, it will respond at an even higher rate (of action potentials) if that central spot is surrounded by a dark ring. This property of an "ON center" and an "OFF surround" means that the neuron signals when there is something bright in the center of its receptive field, but does so even more robustly when that bright "thing" has a dark "thing" right next to it. *Figure 3.3* illustrates how this principle, by pooling across a series of neurons with ON center/OFF surround receptive fields, can yield "feature detectors" that cover adjacent regions of the visual field.

how these facts give us embodied constraints for describing the visual world. With this knowledge, we can now start to address the question *How does the brain accomplish visual perception?* Or, in the parlance of previous sections of this chapter, *What do we mean by*

visual information processing in cortex? We'll start by traveling to Johns Hopkins University in the late 1950s, where a revolutionary set of experiments defined the course of the next half-century of visual neuroscience.

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – BOTTOM-UP FEATURE DETECTION

The V1 neuron as feature detector

The anatomy of projections from retina, via LGN, to the calcarine fissure, combined with the systematicity of visual impairments produced by damage to the calcarine cortex, make it clear that this region plays an important role in visual perception. By 1958, David Hubel (1926–2013) and Torsten Wiesel also knew, from extracellular electrophysiological experiments by themselves and others, that neurons in the LGN responded with a burst of action potentials when a small spot of light was projected in specific regions of the visual field (different regions for neurons receiving input from retinal ganglion cells projecting from different parts of the retina). That is, LGN neurons had spatial receptive fields (see *Thought Box 3.2*). They began studying the functions of primary visual cortex,

therefore, by projecting small spots of light into the eye of a cat with an ophthalmoscope (not unlike the apparatus whose eye pieces you look through when having your vision examined by an eye doctor) while recording from neurons with the extracellular electrophysiological method described in *Chapter 2*. The breakthrough discovery came when, in the process of changing slides in the ophthalmoscope, they found that a neuron responded with bursts of action potentials each time the shadow of the edge of the glass slide passed across the retina at a particular angle (*Figure 3.2*; *Research Spotlight 3.1*; *Web Link 3.1*).

Having made this initial observation, Hubel and Wiesel abandoned tiny spots of light in favor of bars of light (as can be made by cutting a long, narrow slit in a piece of cardboard, then shining a bright light through the slit and onto a projection screen). With this new stimulus, they quickly confirmed that many neurons in V1 responded robustly to a straight-line boundary defined by light on one side and darkness on the other. Importantly, for any given cell, this only occurred when that boundary was oriented in a

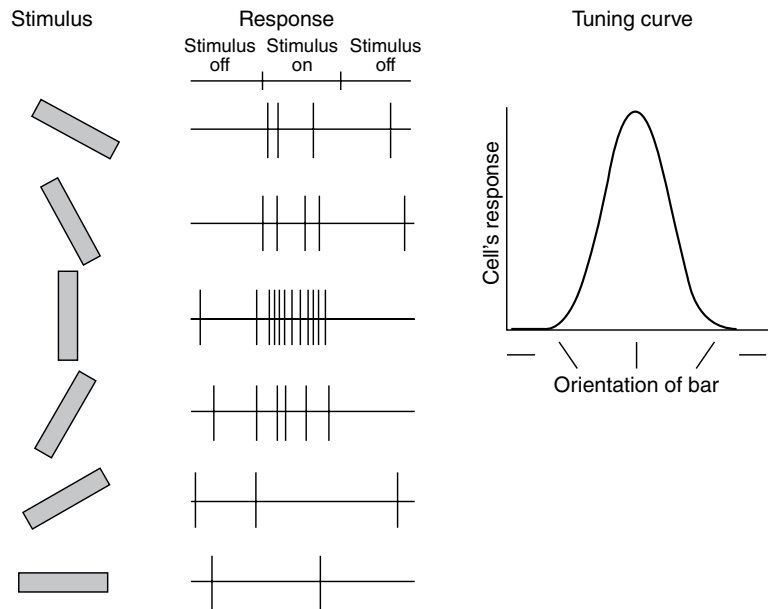


FIGURE 3.2 Orientation selectivity of a simple cell in V1, illustrated by action potentials (“Response” column) generated in response to a bar presented at six orientations (“Stimulus” column), and summarized in a tuning curve. Source: Sekuler and Blake, 1985. Reproduced with permission of Alfred A. Knopf, Inc.

RESEARCH SPOTLIGHT

3.1 Discovery of orientation selectivity in V1

Here's how David Hubel describes this seminal experiment with Torsten Wiesel, the first of many that would lead to the awarding of the Nobel Prize to the duo in 1981. [Some annotations are interleaved, to promote clarity]:

The major breakthrough ... came in our third or fourth experiment. We had isolated a big stable cell [identified a single neuron from which they could obtain reliable recordings with an electrode] which for some hours was unresponsive to anything we did. But as we worked on we began to get vague and inconsistent responses in one region of retina. [They located the spatial receptive field of the neuron] ... We struggled, and seemed to be getting nowhere, when suddenly we started to evoke brisk discharges. We finally realized that the discharges had nothing to do with the dark or light spots but were evoked by the action of inserting the glass slide into the slot. The cell was responding to the faint shadow of the edge of the glass moving across the retina, and it soon became clear that the responses occurred only over a limited range of orientations of the edge, with a sharply determined optimum and no response to orientations more than 30 degrees or so from the optimum. [That is, the neuron only fired a burst of action potentials when the slide was tilted in a particular orientation, and it would no longer respond if the slide were rotated such that the line was tilted more than 30 degrees from the neuron's "preferred" orientation. See *Figure 3.2* for an illustration, and *Web Link 3.1* for a vivid demonstration of such an experiment.]

The importance of this discovery is that it led to a profound conceptual insight about the mechanisms whereby the primary visual cortex begins the computations that result in visual perception. More specifically, as illustrated in *Figure 3.3*, it led to the insight that neurons in primary visual cortex may construct elemental features in the visual scene via the integration of information from the simpler center-surround receptive fields of LGN neurons.

About the discovery itself, Hubel goes on with the following assessment:

This event has sometimes been held up as an example of the importance of "accident" in science. We have never felt that it was an accident. If there is something there to discover one has to take the time to find it, and one has to be relaxed enough about the way one works so as not to foreclose the unexpected. Two other groups failed to discover orientation selectivity because they were too scientific, in a simplistic sense of that word: one group built a device to generate horizontal bright bars, the other group, vertical, in both cases so that they could explore the retina more efficiently than with a roving spot. In a certain early phase of science a degree of sloppiness can be a huge advantage. (Hubel, 1996, pp. 308–309)

particular direction. That is, one neuron might respond preferentially to a boundary oriented vertically, another to a boundary tilted 10° to the right of vertical, another to a boundary oriented horizontally, and so on, such that, with enough recording from enough neurons, they collected evidence that the full 180° of possible orientations would be covered. (Note

that it is only necessary to talk about 180° of possible orientations for a bar because, for example, rotations of 45° and 225° from the same starting point result in the same outcome.) The property of firing action potentials most robustly in response to a narrow range of orientations has come to be known as "orientation selectivity," and is a hallmark property of many V1

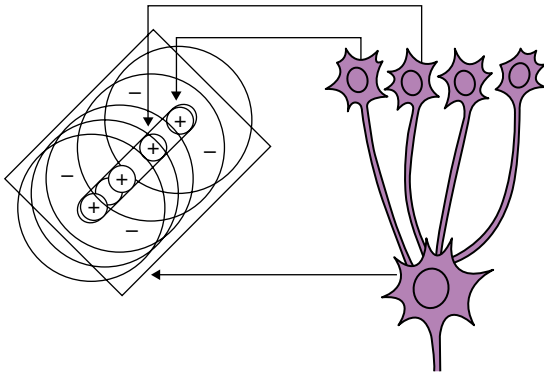


FIGURE 3.3 Constructing orientation selectivity in V1 from LGN neurons with ON center/OFF surround receptive fields. Hubel and Wiesel's idea is the one introduced in *Thought Box 3.2*: the V1 neuron receives input from four LGN neurons whose receptive fields are overlapping, and whose centers are aligned along an imaginary line tilted 45° to the right of vertical. By laying pencil at different orientations across this array of receptive fields, you can prove to yourself that a bar of light oriented at 45° would (1) maximally excite the ON centers of each LGN neuron while (2) minimally stimulating their OFF surrounds. The consequence for the V1 neuron is that this stimulus produces the strongest net input from the LGN, thereby making it (the stimulus) optimal for eliciting a robust response from this V1 neuron. Source: Hubel, 1988. Reproduced with permission of Nature Publishing Group.

neurons. *Figure 3.3* illustrates Hubel and Wiesel's idea for how this property might be constructed by pooling across simpler inputs from the LGN. *Figure 3.2* illustrates how this property can be quantified as a “tuning” function (or “tuning curve”; *Tip 3.8*).

Orientation tuning is a specific example of a broader phenomenon that arises in V1, which is the construction of information (the “inference,” if one were to anthropomorphize) from more elemental information (i.e., the presence or absence of spots of light). Indeed, including orientation, there are at least five other such “transforms” of elemental information into visual features: direction selectivity; higher acuity; disparity between the images projected into the two eyes (i.e., binocular disparity, a type of stereoscopic

Tip 3.8

The term “tuning” curve is an analogy to the concept, from signal processing, of a tuned filter, a circuit used to separate signals in relatively narrow bands of frequencies from signals in wider frequency spectrums. One concrete example is a radio that when tuned, at my house, to 88.7 MHz, passes the frequency carrying Wisconsin Public Radio, and is insensitive to what's being carried on other frequencies, such as the adolescent pap that passes for music, and is preferred by my kids, on *Z-104* (MHz) *FM*.

Tip 3.9

To prove to yourself that the brain is “set up” to infer motion from temporally sequential and spatially offset events, consider the lighted marquee of a theater. A row of light bulbs lines the edges of the marquee, and flashing them rapidly, in succession, creates the illusion of a flash of light “running in circles” around the marquee.

information); the convergence of ON center/OFF surround and OFF center/ON surround (there are both types of LGN neurons), which gives rise to more complex response properties than the one-dimensional orientation-tuned “simple cell” illustrated in *Figure 3.2*; and the construction of a single color axis from three cone types. A detailed consideration of how each of these transforms is computed is beyond the scope of this book. But let's briefly consider just one more example: direction. Imagine receiving information that the receptive field of a neuron in the LGN is responding to visual stimulation at time t , and that an LGN neuron with an immediately adjacent receptive field (let's say, 90° from vertical to the left) subsequently signals visual stimulation at time $t + 1$. Voilà: a neuron receiving such time-lagged inputs from these two cells would “deduce” that an object in this part of the visual field is moving to the left (*Tip 3.9*).

What might be the utility of some of the properties that emerge from processing in V1? Orientation tuning

allows for the detection of edges, a fundamental property of many objects that we encounter in our daily lives. Direction selectivity is a necessary element of motion perception. Stereoscopic information is used to see in depth (i.e., 3-D). Combining inputs from the two classes of LGN receptive fields can produce complex tuning properties such as responding to the end of a line (a property referred to as being “end-stopped”). Finally, with regard to color, it has been estimated that humans can distinguish among millions of hues of color, a feat that surely requires considerable processing of the initial inputs to cortex from just three types of wavelength-sensitive cones. These properties (e.g., edges, motion, binocular disparity, end-stopping, color) can be thought of as low-level visual features that, when combined in different ways, are the constituent elements of the complex objects that we recognize and interact with, seemingly effortlessly, on a daily basis. Thus, an influential way of understanding information processing in V1 is to view these neurons as **feature detectors** – units in the system whose function is to signal the presence of their particular feature each time it appears in their receptive field, and to pass this information forward to downstream networks in the visual system that will aggregate information from many such feature detectors to assemble a representation of what object is currently in the visual field. *Chapter 5* will give more consideration to what we know about object perception, but for now we’ll maintain our focus on the organization of V1.

Columns, hypercolumns, and pinwheels

A common practice during single-electrode extracellular recording sessions is, having characterized the response properties of the first neuron encountered, to advance the electrode further into (or, depending on the angle of penetration, along) the region under investigation. (In later chapters we’ll introduce the technique of simultaneously recording from multiple electrodes.) The results of several perpendicular penetrations (i.e., from layer I to layer VI within a single column), led Hubel and Wiesel to conclude that neurons within a cortical column share the same orientation selectivity. (Hence the term “orientation

column.”) Penetrations that were parallel to the cortical surface, in contrast, revealed that adjacent columns tended to be tuned to orientations that were slightly rotated from those of neighboring columns. This led to the proposal that primary visual cortex was organized into **hypercolumns**, a hypothesized grid of orientation columns that, together, represented every possible orientation that could fall within a receptive field. Of course, to be equipped to analyze every possible feature that could occupy a receptive field, such hypercolumns would additionally need to represent all possible features (or, minimally, the constituent elements needed to construct all possible features) that could fall into this receptive field. Consistent with the idea of omni-featured hypercolumns was the fact that parallel penetrations would periodically encounter columns of neurons that did not show orientation selectivity, but that did seem to process information about color. Thus, Hubel and Wiesel conjectured, each hypercolumn also contained nonselective, chromatically tuned neurons that are important for the perception of color.

A few decades later, advances in functional tissue imaging techniques (*Methodology Box 3.1; Tip 3.10*) led to further specifications, then modifications of the hypercolumn model. It is important to note that, in their studies from the 1950s and 60s, Hubel and Wiesel never directly observed a hypercolumn in toto. Rather, they inferred the existence of this organizing structure from after-the-fact interpretation of the results of tens, if not hundreds, of experimental sessions. An advantage of functional tissue imaging over single-electrode recording is that it can assess the activity of a relatively large region of cortex (on the order of cm^2). A critical early discovery, triggered by the research of Margaret Wong-Riley (1989), was that superficial layers of V1, when assayed to detect activity of the enzyme cytochrome oxidase, revealed “blobs” of tissue that stained more darkly than surrounding tissue. Cytochrome oxidase is a membrane-spanning protein found in mitochondria, and is critical in the generation of adenosine triphosphate (ATP) via oxidative metabolism. Therefore, darker staining, which indicates elevated levels of cytochrome oxidase activity, is a marker for elevated activity in the cells containing it. Subsequently, Margaret Livingstone and David Hubel recorded extracellularly from neurons located within blobs and discovered these to primarily contain

METHODOLOGY BOX

3.1 Functional tissue imaging

In the broadest sense, “imaging” techniques differ from other methods of cognitive neuroscience in that they produce images of the tissue that is being studied. (In contrast, extracellular electrophysiology, for example, produces measurements of the activity of a neuron, but not an image of the neuron itself.) Of the two general categories of imaging – structural and functional – this box introduces three examples of the latter. Within the category of functional imaging, there exists another distinction between those using invasive vs. noninvasive methods. We will use the term “functional tissue imaging” to refer to the former, because each entails performing a craniotomy and taking measurements directly from the tissue of interest. (With noninvasive brain scanning (a.k.a. “functional imaging”), in contrast, one “takes pictures” of tissue and/or activity inside the cranial cavity without having to remove any bone.)

The oldest of the three, 2-deoxyglucose imaging (2DG), entails the systemic administration of glucose molecules that have been tagged with a radioactive ion (such as ^{14}C) and further modified so that they can't undergo glycolysis. It exploits the fact that glycolysis is the primary source of energy for all cells in the body. Once administered, 2DG blends with naturally circulating glucose in the blood stream, and, just like endogenous glucose, is taken up by neurons in an amount proportional to their activity. (Recall, from *Thought Box 2.2*, the considerable energy demands of synaptic activity.) Once taken into a cell, 2DG is trapped. Upon completion of the experiment, the animal is sacrificed, and the tissue of interest dissected out of the brain and placed against a radiographic plate that detects the radiation given off by the 2DG that has been trapped in previously active neurons (a process called autoradiography). Results from a 2DG experiment are illustrated in *Figure 3.5*.

A second functional imaging method uses voltage-sensitive dye. It differs from 2DG in three important ways. First, it is sensitive to fluctuations in membrane voltage. Second, it gives a real-time measure of activity, whereas 2DG requires extensive after-the-fact processing in order to obtain the results from the experiment. Third, voltage sensitive dye can be used to measure cortical activity in several different conditions (e.g., responses to bars of 8 different orientations). 2DG, on the other hand, requires the sacrifice of the animal after just one experimental condition. Thus, to determine the organization of the neural processing of 8 different orientations, one would need to acquire data from (a minimum of) 8 animals, then try to assemble a composite map across these.

Finally, optical imaging of intrinsic signals entails shining light of a particular wavelength on the tissue of interest, and measuring how the make-up of the reflected light changes as a function of experimental condition. The somewhat coy reference to “intrinsic signals” reflects the fact that there remains some uncertainty about all of the factors that produce changes in tissue reflectance – although the majority of the variance in the signal is due to different reflectance properties of oxygenated vs. deoxygenated blood, there may also be contributions from membrane voltage (despite the absence of externally applied dye), and even from the subtle stretching and compressing that neuronal membranes exhibit under different physiological states.

the orientation-nonselective, chromatically tuned neurons described in the previous paragraph. And so, the hypercolumn model was modified such that each contained a color-sensitive blob.

More recently, optical imaging methods have revealed that, instead of being organized into cube-like modules of tissue organized in tidy rows of columns, the functional organization of V1 is more reminiscent of a series of

Tip 3.10

The method used by Wong-Riley, and many who have subsequently studied cytochrome oxidase activity in the brain, is immunohistochemistry, a technique in which tissue is bathed in a solution containing an antibody to the protein of interest. The antibody, which binds to the protein of interest, has been manipulated so that it (and, therefore, the protein to which it is bound) will be visible under a microscope. In laboratory jargon, the tissue has been “stained,” and/or the protein of interest has been “labeled.”

pinwheels, with an orientation-insensitive blob at the center and several “petals” made up of orientation-selective columns of neurons emanating from it (*Figure 3.4*). *Methodology Box 3.1* distinguishes between functional tissue imaging methods employing 2-deoxyglucose (*Figure 3.5*), voltage-sensitive dye, and optical imaging of intrinsic signals, an important contribution to which are the differing reflectance properties of oxygenated (i.e., arterial) vs. deoxygenated (i.e., venous) blood.

Behind the pinwheel

Before leaving this section, a few words about how blobs come to be blobs, and orientation-selective regions come to be orientation-selective. There are two major classes of retinal ganglion cells (RGCs),

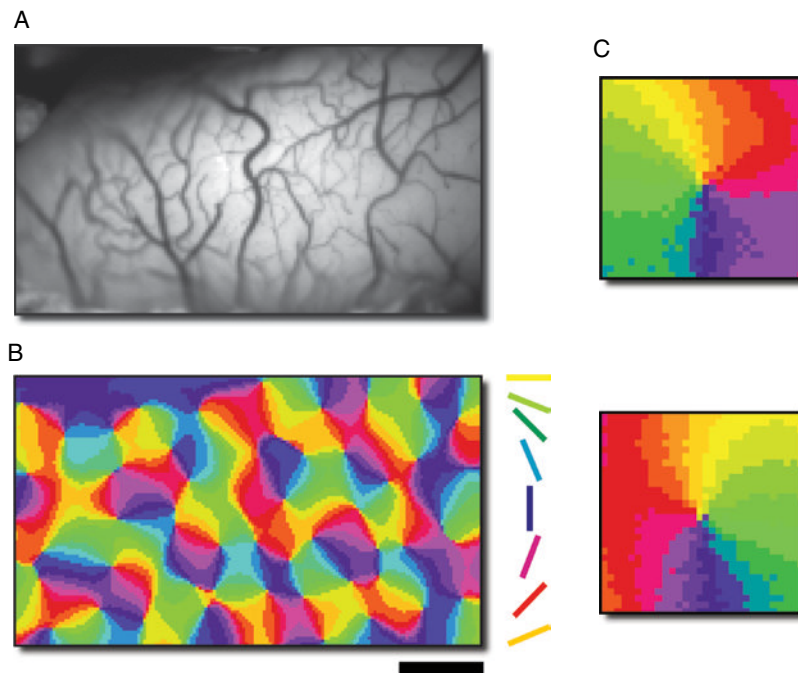


FIGURE 3.4 The organization of visual cortex in the cat, assessed with optical imaging of intrinsic signals. Panel **A** illustrates the surface of the cortex from which the subsequent images were acquired (scale bar = 1 mm). **B**. Map of preferred orientations in the region illustrated in **A**, color-coded to indicate which of the 8 orientations shown on the right-hand side of the figure produced the maximal response at each location. The panels in **C** are close-up images (for these two panels, the scale bar from **B** corresponds to 300 μm) of the centers of two of the pinwheels seen in **B**. Note the smooth transitions from one preferred orientation to the next within each pinwheel. Source: Image courtesy of Tobias Bonhoeffer, based on Bonhoeffer and Grinvald, 1991. Reproduced with permission of Nature Publishing Group.

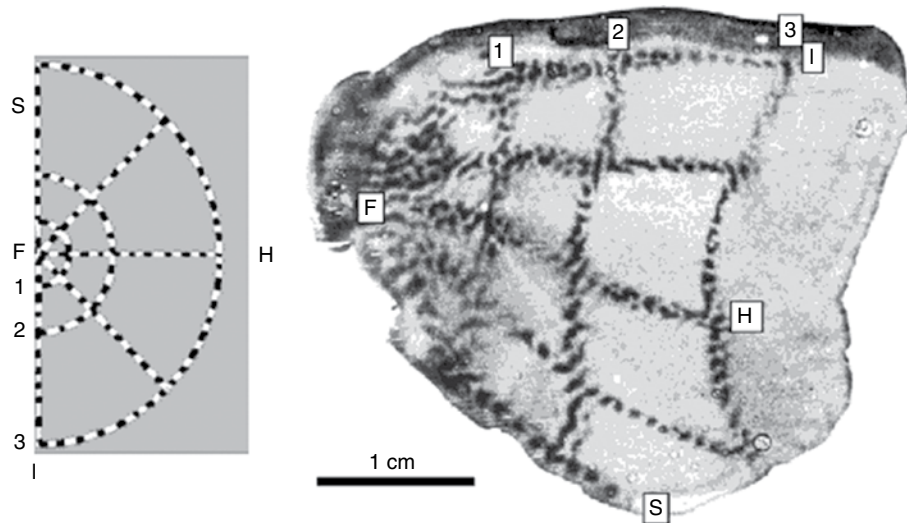


FIGURE 3.5 Retinotopic organization of V1, visualized with ^{14}C 2DG labeling. The stimulus (left side of figure) was presented to one eye of a monkey, with the “F” label indicating the portion projected on to the fovea, “S” and “I” indicating superior and inferior aspects of the vertical meridian, respectively, and “H” indicating the horizontal meridian. Black and white coloring on the lines and concentric circles of the stimulus flickered back-and-forth at a rate of three times per second, to vigorously stimulate activity in V1 neurons into whose receptive fields the stimulus fell. The right of the figure shows an autoradiograph of a section of cortex dissected from primary visual cortex, with the letters (F, S, I, H) indicating the cortical representation of these demarcations of the visual field. Note the consequence of cortical magnification on how the neural representation of the stimulus is “laid out” on the cortex. Source: Tootell, Silverman, Switkes, and De Valois, 1982. Reproduced with permission of AAAS.

that take their names from the *magnocellular* and *parvocellular* neurons of the LGN to which they project. (As is illustrated in *Figure 2.4* and *Figure 5.1*, layers 1 and 2 of the LGN contain magnocellular neurons, and layers 3, 4, 5, and 6 contain parvocellular neurons.) M RGCs have dendritic arbors that span a physically larger surface area, and thus sample from a larger number of photoreceptors than do P RGCs. Two consequences of such broad sampling are that (1) M RGCs have relatively larger receptive fields, and (2) they do not carry color-specific information (whereas P RGCs do). M RGCs also have faster axonal conduction velocities than do P RGCs. It is magnocellular neurons (of the LGN) that project to orientation-selective regions of V1, and parvocellular neurons that project to blobs. This information will be particularly germane in *Chapter 5*, when we consider feedforward projections from V1 to different regions of extrastriate cortex, and how these regions come to have different functional specializations.

INFORMATION PROCESSING IN PRIMARY VISUAL CORTEX – INTERACTIVITY

At this point, the discussion from *Chapter 2* about the projection profiles of different layers of neocortex (*Figure 2.4*) becomes particularly germane. We'll begin with one of the more astounding facts about the anatomy of the visual system; one that, until this point, has been glossed over.

Feedforward and feedback projections of V1

Take a moment to revisit *Figure 2.6*, noting the elegant sweep of the optic radiations that connect the LGN with V1. A new piece of information that will prompt renewed consideration of this tract is that an *estimated 75–90% of the fibers comprising the optic radiations*

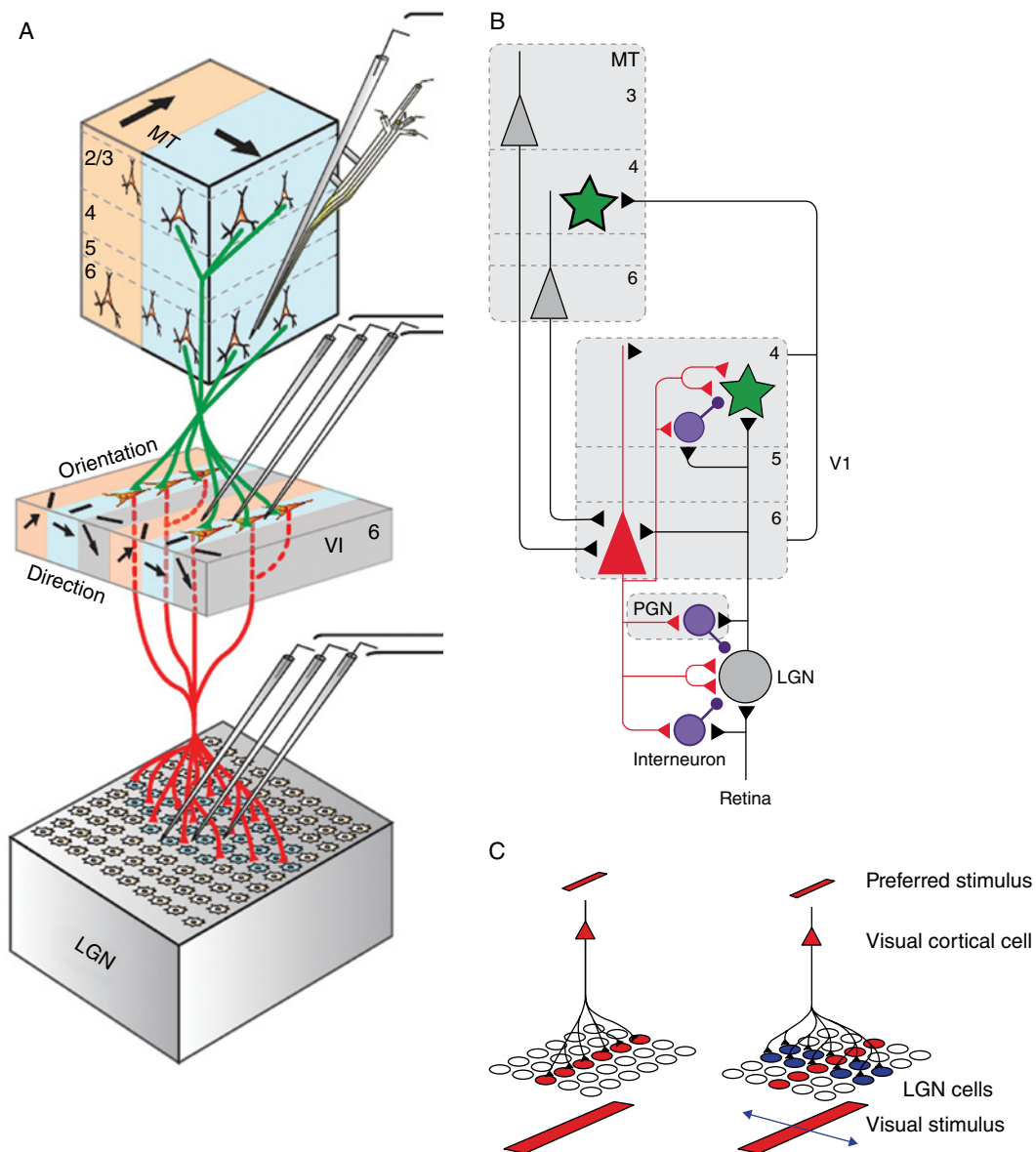


FIGURE 3.6 The role of feedback in determining what we see. **A.** illustrates an experimental set up in which neurons with overlapping receptive fields are simultaneously recorded from in the LGN, layer VI of primary visual cortex (V1), and layer VI of the motion-sensitive region of extrastriate visual cortex known alternately as MT or V5 (we'll use "MT," but see *Tip 3.11*). **B.** Schematic diagram of some of the circuitry that is important for motion perception. The feedforward pathway begins at the bottom of the panel, with an axon from a retinal ganglion cell projecting to and synapsing on a neuron in the LGN of the thalamus, and a branch from this axon synapsing on an adjacent inhibitory interneuron. The LGN sends its "primary" axon to a neuron in layer IV of cortical area V1 (green star), with collaterals to other cell types. The final feedforward step that is illustrated here is from V1 to MT, cartooned as axons projecting from layers VI and IV of V1 to a neuron in layer IV of MT (green star). (Note that this deviates from the "general rule" described in *Chapter 2*, that corticocortical feedforward projections tend to originate from layer II/III. The brain is complicated; one can find an exception to almost all "general rules.") The principle feedback pathway illustrated here starts with projections from MT (layers 3 and 6) to a layer-VI neuron in V1, which,

are traveling from V1 back to the thalamus. That is, for each axon conveying signals from the LGN to V1 (i.e., in the feedforward direction), there are anywhere from three to nine axons conveying feedback signals from V1 to LGN. (This range of estimates can be found in the current literature.) One hint at why our visual systems evolved in this way has come from a series of elegant (and technically very challenging) experiments performed by London (UK)-based neuroscientist Adam Sillito and his colleagues. Their studies, summarized in *Figure 3.6*, provide evidence that ongoing activity in extrastriate and striate cortex can, in effect, shape the signals that the retina, via the LGN, will deliver to them. Cudeiro and Sillito (2006) summarize the more general principle as follows: “For sensory input to be rapidly assessed and to guide the behaviour of an organism, it needs to be continually placed in the context of hypotheses formulated by higher brain centres” (p. 304).

Why does this circuitry require so disproportionate a balance of feedback vs. feedforward projections? One reason is that implementing functions like the anticipatory “wave” of activity in the LGN that is illustrated in *Figure 3.6.C* requires a lot of circuitry. Related to this is the fact that layer VI projections to subcortical structures are much more complicated than what is illustrated in *Figure 3.6.B*. For example, the region labeled “PGN” in this figure, the perigeniculate nucleus, is part of a larger structure known as the thalamic reticular nucleus (TRN; *Figure 2.4*). Although the TRN is present in many of the photographic images in this book that include the thalamus (e.g., *Figure 1.2.B* and *Figure 2.2*), it can’t actually be seen in these figures

Tip 3.11

It has been said of neuroanatomists that “they would sooner use each other’s toothbrushes than use each other’s terminology.” An exaggeration, of course, but the naming convention for the motion-sensitive region of extrastriate cortex illustrates this point. It was discovered more-or-less simultaneously in two laboratories, located on two sides of the Atlantic. The UK-based discoverer named it “V5” (“V”s 1–4 having already been named). The US-based discoverers named it “area MT,” after the middle temporal lobe of the owl monkey in which they discovered it. As a result, in the literature one finds references to “V5,” to “MT,” or to “MT/V5” or “V5/MT.”

due to its microscopic thinness. Unlike most subcortical nuclei, which are clumps of cell bodies whose shape often resembles a familiar object (e.g., an almond, a lens, a tail, ...), the TRN is a thin sheet that wraps around much of the body of the thalamus. (The familiar object that it evokes for me is plastic kitchen wrap.) The name “reticular” derives from the fact that it is reticulated – punctuated with holes necessitated by the fact that hundreds of thousands of axons pass through it while traveling from thalamus to cortex or from cortex to thalamus. The TRN is made up exclusively of GABAergic neurons whose output, in addition to collateral projections within the TRN itself, is exclusively focused on the thalamus. As a result, the TRN is ideally suited to implement an **inhibitory surround**.

FIGURE 3.6 (*Continued*) in turn, projects to the same LGN neuron with which we started (with collaterals to inhibitory interneurons that also synapse on our LGN neuron). (The MT-to-V1 projections also deviate from the “general rule.”) **C.** A schematic “zoom-in” on the LGN, illustrating a functional consequence of this bidirectional network architecture with the example of the processing of information about a bar moving through the visual field (labeled “preferred stimulus”). Depending on the balance of top-down signals from MT to V1, the output of V1-to-LGN feedback projections (schematized as just a single neuron, but in reality coming from a local population of neurons), either emphasize the LGN’s processing of the orientation of the visual stimulus (left side of panel), or its processing of the motion in which the bar is moving (right side of panel). Indeed, in the latter case, corticothalamic feedback signals can produce a “wave” of activity in LGN that precedes the arrival of the stimulus into the receptive field, in effect “preparing” the LGN with a predicted trajectory of the stimulus. Source: Cudeiro and Sillito, 2006. Reproduced with permission of Elsevier.

Corticothalamic loops, center-surround inhibition, and signal-to-noise ratio

To explain the concept of an inhibitory surround, let's return for a moment to the anecdote from *Chapter 2* of the encounter with the bear. This anecdote was first used to illustrate how oscillatory synchrony might contribute to prioritizing the processing of the bear over the bird or the dappling sunlight. Another goal for the brain in this situation, however, would be to generate as sharp a representation of the bear as is possible. One way to accomplish this goal would be to minimize competition from the neural representation of other large-ish, hirsute quadrupeds. (Now would not be the time, for example, to contemplate the similarities and differences of bears to, say, deer, raccoons, and dogs; as in *Hmmm. Is it really true that bears are evolutionarily more closely related to dogs than to cats? ...*) When a cortical neuron involved in the neural representation of the bear (let's call it neuron *A*) sends a volley of action potentials to another neuron in the hypothetical "bear network" (we'll call it *B*), a layer V or VI neuron in the same column as *A* may send a parallel volley of action potentials to a thalamic neuron (*T*) that, itself, is connected with neuron *B*. Thus, *B* receives excitatory (i.e., glutamatergic) inputs from both *A* and *T*. (This is effectively what's illustrated in the feedforward projections from V2 to V4 that are illustrated in *Figure 2.4*.) The real value-added of this corticothalamic circuit, however, is that where the *B*-to-*T* axon passes through the TRN, it sends collateral branches to TRN neurons that, in turn, project to thalamic neurons that surround *T* (but don't synapse with *T* itself). The result is that action potentials from layer VI of cortical column *A* have the effect of inhibiting activity in thalamic neurons surrounding *T* (because excitatory collaterals from the *A*-to-*T* pathway activate TRN neurons that, in turn, release GABA onto these neurons). Back in the cortex, the consequence for *B* (and, by extension, for the neural representation of the bear) is lowered activity in competing neurons (representations), due to a reduction of their excitatory thalamic "drive." A more formal way to summarize this is that engagement of this corticothalamic loop increases the signal-to-noise ratio (SNR) in the information carried by *A* and *B*. The same network dynamics described here (minus the part about the bear) have been demonstrated in studies of

attention, which we will take up in *Chapter 6*. (The more general mechanism of center-surround inhibition will also be central to understanding functional contributions of the basal ganglia to motor control (*Chapter 7* and *Chapter 8*).

And so, to return to the question of why the disproportionate ratio of corticothalamic to thalamocortical fibers in the optic radiations, another part of the answer is the implementation of cortical center-surround control of the LGN. Note that, although the example in the previous paragraph used just one layer V or VI neuron in cortical column *A*, the inhibitory surround is, in fact, implemented by descending fibers from many neurons. A second function implemented by corticothalamic fibers, and therefore also implicated in the makeup of the optic radiations, is controlling the "thalamic wake-up call."

Setting the state of thalamic neurons: tonic mode vs. burst mode

Another function of corticogeniculate fibers is to modulate the effectiveness with which LGN neurons can convey signals to the cortex. They can do this because of a property of thalamic neurons that distinguishes them from most cortical pyramidal neurons: thalamic neurons can display two distinct modes of firing action potentials, tonic and burst mode. Burst-mode spiking has been referred to as "a wake-up call" from thalamus to cortex, for reasons that are evident in *Figure 3.7*: burst-mode action potentials are markedly more effective at transmitting information than are tonic-mode action potentials. Thalamic neurons move into tonic firing mode when their membrane potential is maintained at a depolarized level for 100 ms or longer, and into a burst mode when maintained at a hyperpolarized level for 100 ms or longer. Research from Sillito's group, using a procedure not unlike that illustrated in *Figure 3.6.A*, has established that the layer VI neurons in V1, via projections to GABAergic neurons in the TRN and within the LGN itself, play a role in controlling whether visual information from the LGN is conveyed to V1 via tonic or bursting spike trains. Thus, in effect, the function of some of the fibers descending from cortex to thalamus via the optic radiations enable the cortex to decide when to give itself a wake-up call.

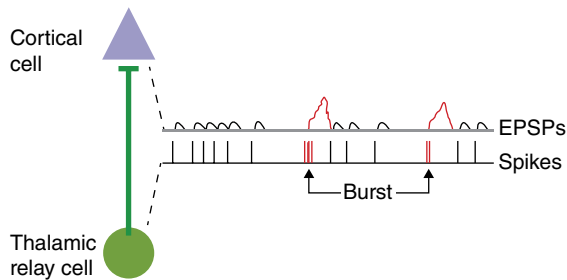


FIGURE 3.7 A schematic interpretation of the effects of thalamic action potentials on layer IV cortical cells. The line labeled “spikes” is a timeline periodically punctuated by action potentials fired by a thalamic relay cell, with tonic-mode spikes illustrated in black, and burst-mode spikes illustrated in red. The corresponding timeline labeled “EPSPs” represents the excitatory postsynaptic potential that each thalamic spike produces in the downstream cortical neuron. The larger cortical response to burst-mode firing results from a larger initial EPSP and the temporal summation of EPSP from the burst-mode spikes, which occur more closely together in time. Source: Sherman, 2001. Reproduced with permission of Nature Publishing Group.

(Although cognitive neuroscientists who study visual perception in humans typically can’t study the visual system with the circuit-level resolution emphasized in much of this chapter, *Web Clip 3.4* illustrates how a detailed understanding of the cellular- and circuit-level functional organization of the early visual system can be leveraged to carry out sophisticated neuroimaging studies of, in this case, 3-D visual motion perception. And to understand the imaging method featured in *Web Clip 3.4* you’ll first want to be familiar with the concepts covered in *Web Clip 3.1*, *Web Clip 3.2*, and *Web Clip 3.3*.)

Circularity? It can depend on your perspective

On the surface, some of the phrases used in the previous two sections, such as that a “function of [layer VI of cortex] is to modulate the effectiveness with which LGN neurons can convey signals to the cortex” can sound circular. Indeed, it was physically difficult for me to type that sentence – some part of my motor

system “knew” that it didn’t sound right. When given more consideration, however, one can see that what may make one-sentence summaries of findings like these seem paradoxical is the absence of context. To interpret these findings on their own illustrates that it can be a thorny proposition to attempt to isolate discrete “processes” (or “functions”) within the spectacularly complex, nonlinear system that is the brain. For example, in defining a process, where is one to stop? Take the case of the proximal control that V1 exerts over whether LGN activity emphasizes the static orientation of a stimulus or its motion (*Figure 3.6.C*). Whereas this could be characterized as an instance of feedback control of V1 over the LGN, we saw that this network is embedded in a larger network that includes MT, with inputs from the latter influencing layer VI of V1. And, although we won’t get there until *Chapter 5*, the reader won’t be surprised to learn that there are analogous networks operating at an even larger scales, and using more abstract information (e.g., the behavioral context) to influence whether MT should influence V1 to influence the LGN as to whether orientation or direction information should be prioritized. Thus, the “local” examples of feedforward/feedback interactions that we’ve considered here are a part of the active “churning” of the brain that was invoked in the *Introduction to Section II*. Cudeiro and Sillito (2006) make this point explicitly by invoking the “[continual placing] in the context of hypotheses.” We will return to these ideas at the conclusion of this chapter, but first, an example of physiological “churning” that operates over the scale of the entire visual system.

The relation between visual processing and the brain’s physiological state

In the 1920s, a German physician and physiologist named Hans Berger (1873–1941) began studying the electrical properties of the brain via recordings made with an electrode affixed to the scalp (a technique known as electroencephalography (EEG); see *Methodology Box 3.2*; *Figure 3.8* and *Figure 3.9*; and *Tip 3.12* and *Tip 3.13*). In a 1929 publication that was widely dismissed in his native country, he reported that the EEG was strongest when recorded over the

METHODOLOGY BOX

3.2 Electroencephalography (EEG)

Figure 3.8 (adapted from Luck, 2005) illustrates principles of EEG generation. **A.** Depolarization of apical dendrites via the release of glutamate (from “Axons from Presynaptic Cells”), results from the inward flow of positively charged ions, producing a negative extracellular charge. The release of GABA by interneurons at basal dendrites and onto the cell body has the opposing effect, the net result being that the neuron acts as a small **dipole** (*Tip 3.12*). **B.** Cartoon of a cross-section of cortex with a sulcus in the middle. Incoming neural signals to the neurons in the gyrus and sulcal wall on the left (arrows) turn these neurons into dipoles, producing the distribution of charges illustrated with “–” and “+” symbols on the exterior and interior surface of the cortex, respectively. **C.** The summation of all the (thousands, if not more) tiny dipoles illustrated in **B** produces the distribution of charges on the scalp that are measured with EEG electrodes. Although in this example the true source of the currents measured at the scalp is the thousands (if not more) of neurons illustrated in **B**, the identical distribution of charges observed at the scalp could have been generated by the single “equivalent current dipole” illustrated in **C**. Indeed, the physics of the summation of electrical charges from multiple sources is such that, when a pattern of charges is observed on the surface of a sphere (i.e., on the scalp), there is an infinite number of configurations of dipoles inside the sphere that could have produced that pattern. This is known as the **inverse problem**, and its nature is such that it can never be solved with certainty. Nonetheless, there are algorithmic approaches that estimate “solutions” to the inverse problem by making a variety assumptions about which of the infinite number of possible “source solutions” is the most likely for any particular data set.

There’s a sense in which the EEG signal measured at the scalp is a kind of “macro LFP,” in that it reflects the summation of many, many local cortical LFPs. This is necessarily the case, due to the conductive properties of all the tissue lying between cortex and scalp. Note that whereas the changes in membrane potential of a single neuron varies within a range spanning roughly 100 millivolts (e.g., from –90 to +20 mV), the LFP, which is recorded from extracellular space adjacent to groups of neurons typically varies within a range of roughly 10 μ V (i.e., 2-to-3 orders of magnitude smaller). As measured at the scalp, a cortically derived electrical signal needs additionally to pass through blood vessels, CSF, meninges, skull, muscles, and skin, all of which attenuate and blur it. This has at least three implications for the EEG. First, to be detectable, the EEG necessarily reflects the summation of signals from tens of thousands (or more) neurons. Second, the fluctuations in the membrane potential of these neurons needs to be synchronized – otherwise, negative and positive potentials would cancel each other out. Finally, the EEG signal emphasizes signal from superficial layers of cortex, with smaller contribution from deeper layers and negligible contributions from subcortical structures. *Web Clip 3.5* illustrates the collection of EEG data, and *Web Clip 3.6* its analysis.

occipital cortex, and that it had two key properties. First, it was characterized by “continuous oscillations in which ... one can distinguish larger first order waves with an average duration [i.e., cycle] of 90 milliseconds and smaller second order waves of an average duration of 35 milliseconds.” Thus, he was describing

a complex wave form that comprised the summation of oscillations at two different frequencies. (See *Figure 2.10* for analogous properties of the local field potential (LFP); *Figure 4.3* for the separation and measurement of different frequency components of the EEG signal.) He named the more prominent rhythm *alpha*

and the “second order” rhythm *beta*. The second key property that Berger described was that the alpha rhythm of the EEG was considerably larger when the subject’s eyes were closed than when they were open

Tip 3.12

Note that when enough positive charge migrates along dendrites to depolarize the cell body, as must happen to trigger an action potential, the polarity of this dipole reverses. Such reversals of polarity contribute to oscillations recorded in the LFP and in the EEG.

Tip 3.13

Berger’s is a fascinating story that could be the basis of a compelling screenplay or historical novel. (Perhaps it already has been?) As a younger man, while serving as a cavalry officer in the Prussian army, he sustained an injury when he fell off a horse. During his convalescence he learned that his sister, situated hundreds of miles away, had had a premonition that ill fortune had befallen her brother on the very day that he had, in fact, sustained his injury. Years later, as the director of a psychiatric institute, he devised his EEG experiments, and carried them out in a clandestine laboratory, in the hopes of finding a physical/physiological source for the extrasensory perception that he was convinced that his sister had experienced.

(*Figure 3.10*). This fact can be interpreted in at least two ways. One interpretation is that brain systems (like the visual system) may slip into default “idling” states when they are not actively processing information. From this perspective, the opening of the eyes, and consequent delivery of visual information to the cortex, gives the visual system “something to do,” and thereby pushes it out of its idling mode. (Refer to *Methodology Box 3.2* for the idea that a decrease in alpha-band power likely reflects a desynchronization of some of the thousands of neural elements that were previously oscillating synchronously (*Tip 3.15*).) A second interpretation is that the brain may configure itself in different states to optimize performance on the currently prioritized task. Visual perception is

Tip 3.14

Berger’s findings didn’t gain widespread acceptance until they were subsequently replicated in 1934 by the British neurophysiologist Edgar Adrian, who had been awarded the Nobel Prize in Physiology or Medicine in 1932. In addition to replicating the eyes-open/eyes-closed effect originally reported by Berger (and illustrated in *Figure 3.10*), Adrian’s control experiments included a demonstration that the magnitude of alpha-band oscillations remained high with the eyes open, but the subject’s head inside a dark box, indicating that it’s the absence of visual information, not the position of the eyelids per se, to which alpha-band dynamics correspond.

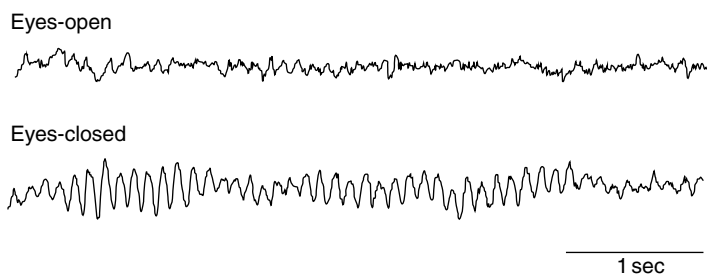


FIGURE 3.10 EEG traces illustrating the “Berger rhythm,” recorded from channel O_1 while the subject sits quietly with eyes open (top row) and closed (bottom row). Note how, in the eyes-closed trace, the magnitude of the alpha-band oscillation fluctuates over time. Source: Data collected by Dr O. Gosseries in the author’s laboratory.

often (although not always) prioritized when the eyes are open. We shall also see, in *Chapter 6*, that the dynamics of alpha-band oscillations vary systematically with the control of visual attention. Regardless of the function(s) that one may want to ascribe to alpha-band oscillations, *Research Spotlight 3.2* describes an objective demonstration that the magnitude of

alpha-band oscillations at any moment in time influences the physiological property of cortical excitability. However, before we can dive into this experiment, we must understand the investigational tools used to obtain the data. *Methodology Box 3.3* will familiarize us with the method of transcranial magnetic stimulation (TMS).

METHODOLOGY BOX

3.3 Transcranial magnetic stimulation (TMS)

Transcranial magnetic stimulation works on the principle of electromagnetic induction – the same principle at work in the turbines at electric power generation plants, or in the pickup of an electric guitar. In secondary school physics, many of us learned this as the “right-hand rule” – if you curl the fingers of your right hand and extend your thumb, current flowing in the direction that your thumb is pointing will generate a magnetic field oriented in the direction of your fingers. Conversely, electrical current will be induced in a length of conductive wire (pointing in the same orientation as your thumb) experiencing a magnetic field (in the same orientation as your curled fingers). With TMS, the TMS coil produces a magnetic field that, when held near the scalp, is experienced by fibers in the brain (*Figure 3.11*). Because these fibers have the properties of conductive wire, the magnetic pulse induces a “pulse” of current in these fibers. The biophysics of cortical tissue is such that current induction is maximal at places where there are geometric changes in the fibers’ pathways (such as where efferent axons bend at right angles to plunge into white matter, or where afferent axons bend at right angles upon emerging from white matter to begin running parallel to the cortical surface) and where there are discontinuities in cell membranes, such as at the axon hillock and at synaptic terminations. Thus, the effects of TMS on a region of cortex are decidedly nonspecific, in that it induces current in incoming fibers, in local circuitry, and in fibers exiting the region.

The effects of TMS on the brain depend on strength of the pulse, with moderately strong pulses capable of generating action potentials not only in neurons that directly experience the magnetic flux of the TMS, but also in neurons that are one or more synapses downstream from the stimulated tissue. For example, in a manner analogous to the effects produced by Fritsch and Hitzig (1870) in their studies of direct microstimulation of motor cortex (see *Chapter 1*), single pulses of TMS delivered to the motor cortex of humans can produce a twitch in a finger or thumb. It is certain that the magnetic field generated by a pulse of TMS does not physically reach the muscles of the hand. Thus, the only explanation for the systematic production of muscle twitches is that TMS of the motor cortex generates action potentials that travel from brain to spinal cord, and are sufficiently robust to trigger action potentials in spinal motor neurons (see *Chapter 7*) which, in turn, are the proximal cause of contraction of muscles in the hand.

TMS is considered a noninvasive tool for the study of brain function, because it entails neither surgical intervention nor lasting damage to the stimulated tissue. Although the technique often elicits questions about safety (e.g., *Can it elicit seizures?*), the method has a lengthy and well-documented history of safe application in cognitive neuroscience research (see, for example, the report by Rossi et al. (2009) that is listed in *Further Reading*).

Methodology Box 3.3 *Continued*

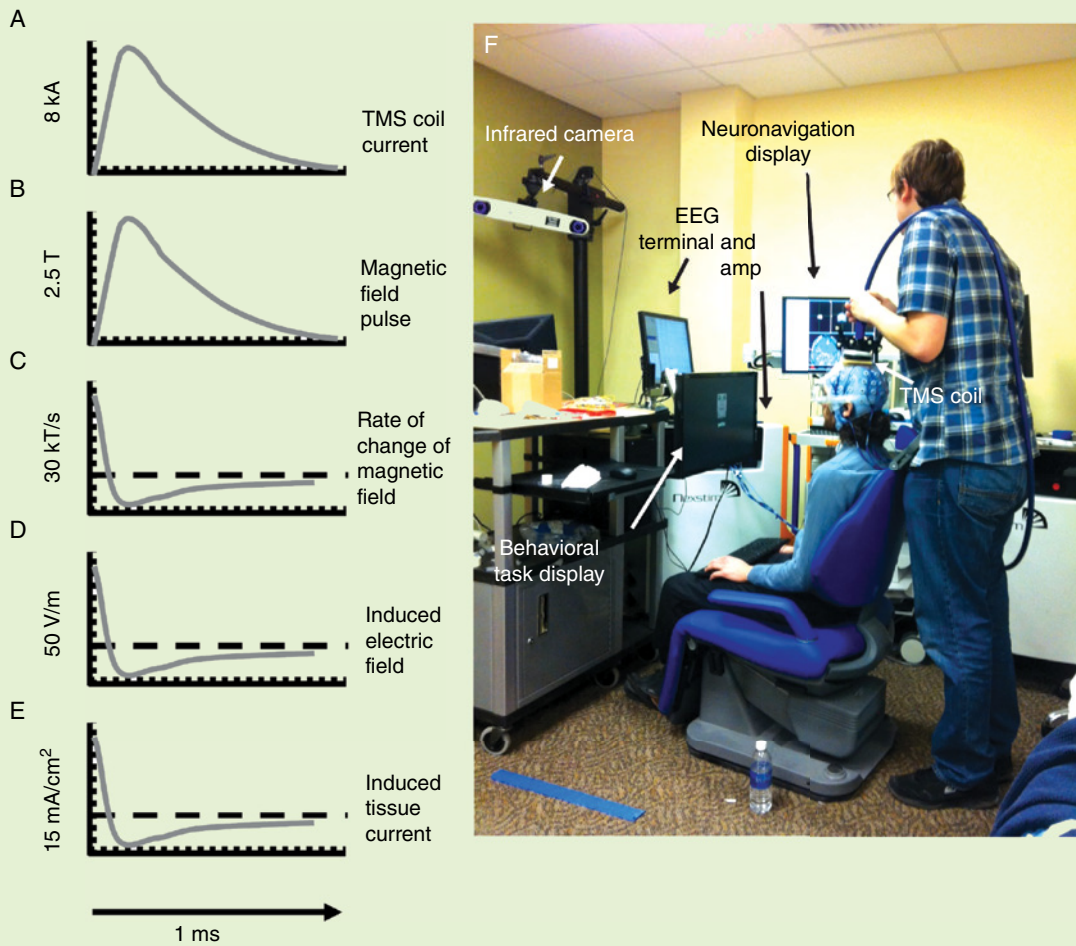


FIGURE 3.11 Transcranial magnetic stimulation (TMS). **A.** The TMS stimulator – essentially a bank of capacitors – discharges a pulse of current through the TMS coil. (The properties illustrated in panels **A–E** are for a circular coil, although different configurations, including two-lobed “figure-of-8” coils, are also commonly used.) **B.** The time course of the resultant magnetic field that emanates from TMS coil (and that is experienced by the brain) mirrors instantaneously that of the current pulse. **C–E.** The time course of the rate of change of the magnetic field, the induced electric field, and the induced current in tissue. These illustrate that the current-induction produced by a pulse of TMS effectively lasts for roughly 300 microseconds or less. The magnitude of each of these effects varies with the amount of current released by the stimulator, a parameter often indexed in terms of “percentage of maximum stimulator output.” **F.** An experimental set-up in the author’s laboratory. The subject is seated and viewing stimuli on the behavioral-task display screen (responding via keyboard in his lab), and is wearing an EEG electrode cap. On his face, and on the TMS coil, are infrared-reflective beacons that are detected by the infrared camera, and used to coregister the location of the TMS coil and the subject’s brain, so as to compute a real-time estimate of where TMS current induction will be maximal in his brain. The experimenter uses this real-time display for “neuronavigation” – positioning the coil so as to precisely target the intended region of the brain. (The TMS stimulator, located to the right of the subject and experimenter and therefore obscured from view in this photo, is roughly the same size as the housing enclosing the EEG amplifier.) Source: Panels **A–E** courtesy of Matthew Rushworth, adapted from Walsh and Cowey, 2000. Reproduced with permission of Nature Publishing Group. **F** is a photograph of the author’s laboratory.

RESEARCH SPOTLIGHT

3.2 Assessing fluctuations in cortical excitability with TMS-evoked phosphenes

The inverse relationship between the magnitude of alpha-band oscillations recorded at occipital EEG sites and the efficacy visual processing, illustrated in *Figure 3.10*, is also seen with spontaneous fluctuations in the magnitude of the ongoing alpha rhythm. For example, the size and latency of cortical responses varies systematically with these fluctuations, as does the likelihood that an individual will be able to detect the presentation of a faint stimulus. This has led to the idea that alpha-band oscillations may produce a state in which the cortex is less able to process sensory information, perhaps because it is “locked in” to this stereotyped, internally driven rhythm. In the study spotlighted here, Romei and colleagues tested the idea that cortical networks are in a less excitable state when the magnitude of alpha-band oscillations is high, vs. when it is low. Lower excitability in a region of cortex would be analogous to hyperpolarization in an individual neuron – in such a state it would be less likely to transmit signals to other brain areas.

Romei et al. (2008) took advantage of the fact that single pulses of TMS delivered to occipital cortex can produce a perceptual phenomenon known as a phosphene – a cloud-like flash of light that a subject perceives even when the eyes are closed and covered with a blindfold (*Fig. 3.12.B*). Phosphenes are presumed to be generated by the propagation of action potentials through the visual cortex. The generation of phosphenes with TMS is probabilistic, with the probability that any given pulse of TMS will generate a phosphene increasing with the magnitude of the TMS pulse. The procedure of Romei et al. (2008) was quite simple, in that they determined for each subject the intensity of TMS that would produce a phosphene on 50% of trials, then delivered many pulses at that intensity. After each pulse they asked the subject to report whether or not s/he experienced a phosphene, and sorted trials accordingly. A final detail is that the EEG was being recorded for each subject right up until the time that TMS was delivered. Thus, upon completion of an experiment, the researchers could determine whether the EEG immediately prior to the delivery of TMS differed systematically on trials when TMS did vs. did not produce a phosphene.

As illustrated in top and bottom panels of *Fig. 3.12.C*, there was, indeed, a systematic difference in the EEG that “predicted” whether or not the subject would experience a phosphene on each trial (*Tip 3.16*): on average, the magnitude of alpha-band oscillations was lower on trials when subjects did detect a phosphene than on trials when they did not. From this, the authors concluded that “momentary levels of posterior alpha-band activity [reflect ...] distinct states of visual cortex excitability, and [... this] spontaneous fluctuation constitutes a visual operation mode that is activated automatically even without retinal input” (p. 2010). This study thus reinforces the idea, built up through the second half of this chapter, that visual perception depends, in part, on the internally determined state of the visual system; a state that exists prior to the initial delivery of information from the outside world (see *Web Clip 3.7*).

Research Spotlight 3.2 Continued

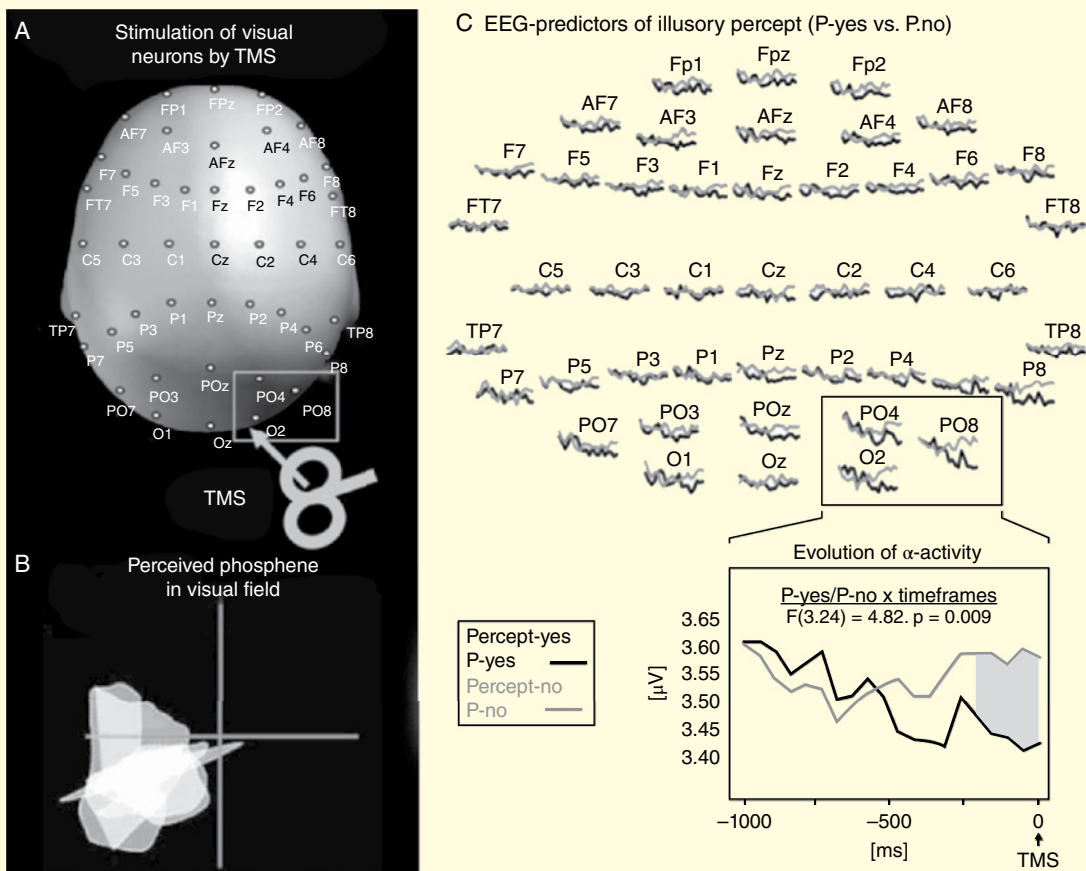


FIGURE 3.12 Methods and results from Romei et al. (2008; see *Web Clip 3.7*). Panel **A** illustrates the delivery of TMS to a region of the scalp medial to electrode O2, thus targeting a dorsal portion of visual cortex. (Small circles and labels correspond to EEG electrodes.) **B**. An “artist’s rendering” of the shape and location of TMS-evoked phosphenes reported by subjects in this study. Note that the location in the visual field corresponds to the known functional organization of the early visual system. Panel **C** illustrates the temporal evolution in the magnitude of alpha-band oscillations across the 1000 ms prior to TMS delivery on trials when subjects did (“Percept-yes”) vs. did not (“Percept-no”) report perception of a phosphene. The upper portion of this panel shows these measures at each of the electrodes illustrated in Panel **A**; the lower portion of this panel shows the average of these measures from the three electrodes at which the alpha-band magnitude for the roughly 250 ms prior to TMS delivery predicted phosphene perception (*Tip 3.16*). Source: Adapted from Romei, Brodbeck, Michel, Amedi, Pascual-Leone, and Thut, 2008. Reproduced with permission of Oxford University Press.

Tip 3.15

Due to the importance of the synchronization of neural activity for the EEG, the reduction in amplitude of an oscillation within a particular frequency band of the EEG is often referred to as a “desynchronization” within that frequency band. Thus, for example, the opening of previously closed eyes is said to produce desynchronization in alpha-band oscillations recorded at occipital electrode sites.

Tip 3.16

In cognitive neuroscience, the word “predict” is often used a retrospective sense, in that the data being analyzed all correspond to events that literally occurred at some time in the past (i.e., data are necessarily analyzed after they have been collected). Nonetheless, if event A occurred prior to event B, and the two are correlated, it is often said that “A predicts B.” (Indeed, temporal separation of the events is not always necessary, as we shall see, for example, in the *Chapter 8* discussion of multivariate pattern classification.)

WHERE DOES SENSATION END? WHERE DOES PERCEPTION BEGIN?

We started this chapter by describing the visual system from what I’ve described as the intuitive, feedforward perspective: photons impact the retina and are transduced into a neural signal; retinal ganglion cells carry this neural signal to the LGN; the LGN relays this neural signal to V1, where features are detected (or, if you prefer, constructed). At one level of analysis, this is necessarily accurate: it’s well-established that damage to the retina and to various locations along the feedforward pathway from retina to cortex produces frank blindness. Additionally, in *Chapter 5* we will see how regions at higher stations of the visual system (including V4 and MT, plus many regions further downstream) construct image and object representations

from the features first detected in V1. That is, extrastriate regions accomplish perception with the features first detected in V1. From this perspective, then, one may want to draw the sensation–perception boundary for vision at V1. Complicating this clear-cut, tidy demarcation, however, is the influence of feedback projections from higher regions on the functions of V1 and the LGN. That is, we have seen that many neural operations that, from one perspective, are very clearly associated with sensation, are nonetheless heavily influenced by concurrent *perceptual* processing at higher levels of the system. This complicates the labeling of any given stage of processing as “belonging” to sensation or to perception. (And we shall see, in *Chapter 5* and *Chapter 9*, that top-down influences from long-term memory influence perception in a manner analogous to how we’ve just documented top-down influences of perception on sensation.) These are hallmarks of a complex system, the idea introduced at the end of *Chapter 2*.

The section of this chapter on *interactivity* described two experiments illustrating how the cortex can exert control over how the thalamus delivers sensory information to the cortex. These two examples can be seen as concrete implementations of the principle of **predictive coding** – the idea that a core function of the brain is to construct, and constantly update, a model of what’s likely to happen in the future. (Such a model is what Cudeiro and Sillito (2006) are referring to as “hypotheses formulated by higher brain centres” (*Tip 3.17*).) We have already discussed how this principle can be seen to play out across multiple spatial scales (i.e., LGN–V1 bidirectional interactions; LGN–V1–extrastriate bidirectional interactions; fluctuations in the alpha-band of the EEG recorded over occipital cortex). This principle can also play out across multiple time scales. In the example of the anticipatory wave of activity across the LGN described in *Figure 3.6.C*, the prediction is about where a moving stimulus will be in the next few milliseconds. In an example of, say, how the author will choose to invest whatever proceeds he may earn from sales of this book, the prediction may be about what he thinks his life will be like after retirement.

Just as the boundaries between sensation and perception and between perception and memory are blurry ones, so too are the boundaries between scientific

Tip 3.17

Harkening back to the continuum of localization-vs.-distributed processing that was introduced in *Chapter 1*, some might quibble that the “hypotheses” that Cudeiro and Sillito (2006) invoke can’t be localized to “higher centres,” but, instead, are emergent from the workings of the entirety of the nervous system.

disciplines devoted to the study of the brain. The principle of predictive coding falls squarely under the rubric of cognitive neuroscience, and has been very useful for understanding the results of experiments such as those that we have reviewed in this chapter, and many, many more that are to come in subsequent sections of this book. At a blurry boundary along the continuum between cognitive neuroscience and theoretical neuroscience there are several theories that have

been proposed that situate the principle of predictive coding within broader models of brain function. One example is theories framed in terms of dynamical systems, a branch of mathematics, in which brain function can be portrayed as a continual effort to minimize entropy. Another is theories framed in terms of Bayesian statistics, whereby the brain’s predictions about the world are formalized as a set of prior probabilities that are periodically updated with information from the environment. Although much of cognitive neuroscience is carried out at more specific levels of analysis, on more specific aspects of human behavior (many of which are captured by the titles of chapters of this textbook), it can be useful while pursuing this research to be mindful of how it relates to “big-picture” theories such as these. For the present, we can keep the principle of predictive coding in the back of our minds as we proceed, in *Chapter 4*, to explore concrete principles of sensation and perception in sensory modalities other than vision.

END-OF-CHAPTER QUESTIONS

1. What is the principle of retinotopic organization? What is cortical magnification and why does it occur?
2. What is a receptive field? How does the size of receptive fields vary as one moves from the fovea to the periphery? As one moves from retina to LGN to V1?
3. In what sense do neurons in V1 act as feature detectors? What kinds of features are represented in V1? These features are not explicitly represented at the level of the retina or the LGN, so how do they come to be represented in V1?
4. Any of the features invoked in Question 3 can be present at any location in the visual field. How is this reflected in the “macro-level” organization of V1?
5. Provide evidence of the statement that *the cortex can exert control over how the thalamus delivers sensory information to the cortex*.
6. What functions are supported by feedback projections from V1 to the LGN? How, at the level of network mechanisms, are they implemented?
7. What factors influence the electroencephalogram (EEG) signal as it is measured at the scalp? What neurophysiological processes can be inferred from the EEG?
8. What is the phenomenon of alpha-band oscillations in the EEG? How do dynamics in the magnitude of these oscillations relate to visual processing?
9. What principle of physics underlies transcranial magnetic stimulation (TMS)? What are some limitations of TMS (think about precision/selectivity and about what areas of the brain it can vs. cannot reach directly)?
10. What is the principle of predictive coding? What evidence is there that this principle may underlie many aspects of the functioning of the visual system that were considered in this chapter?

REFERENCES

- Cudeiro, Javier, and Adam M. Sillito. 2006. "Looking Back: Corticothalamic Feedback and Early Visual Processing." *Trends in Neurosciences* 29 (6): 298–306. doi: 10.1016/j.tins.2006.05.002.
- Fritsch, Gustav, and Eduard Hitzig. 1870. "Über die elektrische Erregbarkeit des Grosshirns." *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin* 37: 300–332. [Available in English translation as: Fritsch, Gustav, and Eduard Hitzig. 2009. "Electric excitability of the cerebrum (Über die elektrische Erregbarkeit des Grosshirns)." *Epilepsy and Behavior* 15 (2): 123–130. doi: 10.1016/j.yebeh.2009.03.001.]
- Gross, Charles G. 1998. *Brain, Vision, Memory*, Cambridge, MA: MIT Press.
- Hubel, David H. 1996. "David H. Hubel." In volume 1 of *The History of Neuroscience in Autobiography*, edited by Larry R. Squire, 294–317. Washington, DC: Society for Neuroscience. <http://www.sfn.org/about/history-of-neuroscience/~media/SfN/Documents/Autobiographies/c9.ashx>.
- Luck, Steven J. 2005. *An Introduction to the Event-related Potential Technique*. Cambridge, MA: MIT Press.
- Malmivuo, Jaakko, and Robert Plonsey. 1995. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. New York: Oxford University Press.
- Romei, Vincenzo, Verena Brodbeck, Christoph Michel, Amir Amedi, Alvaro Pascual-Leone, and Gregor Thut. 2008. "Spontaneous Fluctuations in Posterior [alpha]-Band EEG Activity Reflect Variability in Excitability of Human Visual Areas." *Cerebral Cortex* 18 (9): 2010–2018. doi: 10.1093/cercor/bhm229.
- Sekuler, Robert, and Randolph Blake. 1985. *Perception*. New York: Alfred A. Knopf.
- Sherman, S. Murray. 2001. "A Wake-Up Call from the Thalamus." *Nature Neuroscience* 4 (4): 344–346. doi: 10.1038/85973.
- Wolfe, Jeremy M., Keith R. Kluender, Dennis M. Levi, Linda M. Bartoshuk, Rachel S. Herz, Roberta Klatzky, ... Daniel M. Merfeld. 2012. *Sensation and Perception*, 3rd ed. Sunderland, MA: Sinauer Associates.
- Wong-Riley, Margaret T.T. 1989. "Cytochrome Oxidase: An Endogenous Metabolic Marker for Neuronal Activity." *Trends in Neurosciences* 12 (3): 94–101. doi: 10.1016/0166-2236(89)90165-3.

OTHER SOURCES USED

- Adrian, Edgar D. 1934. "Electrical Activity of the Nervous System." *Archives of Neurology and Psychiatry* 32 (6): 1125–1136. doi: 10.1001/archneurpsyc.1934.02250120002001.
- Blasdel, Gary G., and Guy Salama. 1986. "Voltage-Sensitive Dyes Reveal a Modular Organization in Monkey Striate Cortex." *Nature* 321 (6070): 579–585. doi: 10.1038/321579a0.
- Bonhoeffer, Tobias, and Amiram Grinvald. 1991. "Iso-Orientation Domains in Cat Visual Cortex Are Arranged in Pinwheel-Like Patterns." *Nature* 353 (6343): 429–431. doi: 10.1038/353429a0.
- Grinvald, Amiram, Edmund Lieke, Ron D. Frostig, Charles D. Gilbert, and Torsten N. Wiesel. 1986. "Functional Architecture of Cortex Revealed by Optical Imaging of Intrinsic Signals." *Nature* 324 (6095): 361–364. doi: 10.1038/324361a0.
- Hirsch, Sven, Johannes Reichold, Matthias Schneider, Gábor Székely, and Bruno Weber. 2012. "Topology and Hemodynamics of the Cortical Cerebrovascular System." *Journal of Cerebral Blood Flow and Metabolism* 32 (6): 952–967. doi: 10.1038/jcbfm.2012.39.
- Sillito, Adam M., Javier Cudeiro, and Helen E. Jones. 2006. "Always Returning: Feedback and Sensory Processing in Visual Cortex and Thalamus." *Trends in Neurosciences* 29 (6): 307–316. doi: 10.1016/j.tins.2006.05.001.
- Tootell, Roger B. H., Martin S. Silverman, Eugene Switkes, and Russell L. De Valois. 1982. "Deoxyglucose Analysis of Retinotopic Organization in Primate Striate Cortex." *Science* 218 (4575): 902–904. doi: 10.1126/science.7134981.
- Walsh, Vincent, and Alan Cowey. 2000. "Transcranial Magnetic Stimulation and Cognitive Neuroscience." *Nature Reviews Neuroscience* 1 (1): 73–80.
- Wang, Wei, Helen E. Jones, Ian M. Andolina, Thomas E. Salt, and Adam M. Sillito. 2006. "Functional Alignment of Feedback Effects from Visual Cortex to Thalamus." *Nature Neuroscience* 9 (10): 1330–1336. doi: 10.1038/nn1768.

FURTHER READING

Friston, Karl. 2010. "The Free-Energy Principle: A Unified Brain Theory." *Nature Reviews Neuroscience* 11 (2): 127–138. doi: 10.1038/nrn2787.

A readable summary of many models from theoretical neuroscience addressing high-level questions of what the brain does and why it does it that way.

Glickstein, Mitchell, and David Whitteridge. 1987. "Tatsuji Inouye and the Mapping of the Visual Fields on the Human Cerebral Cortex." *Trends in Neurosciences* 10 (9): 350–353. doi: 10.1016/0166-2236(87)90066-X.

A brief history of the identification and mapping of human visual cortex, including "the brilliant contributions of a young Japanese ophthalmologist, Tatsuji Inouye, which have gone largely unrecognized."

Hubel, David H. 1988. *Eye, Brain, and Vision*. New York: Scientific American Library; Distributed by W. H. Freeman.

A summary of principles of visual neuroscience from one of the field's pioneering researchers.

Luck, Steven J. 2005. *An Introduction to the Event-related Potential Technique*. Cambridge, MA: MIT Press.

Although this book emphasizes event-related potentials (ERP), which are generated by the averaging of EEG signals that are time-locked to events of interest within trials, rather than EEG per se, it is nonetheless an excellent introduction to how EEG data are collected and processed. (Additionally, ERPs will be directly relevant for this textbook, beginning with Chapter 4.)

Rossi, Simone, Mark Hallett, Paolo M. Rossini, Alvaro Pascual-Leone, and The Safety of TMS Consensus Group. 2009. "Safety, Ethical Considerations, and Application Guidelines for the Use of Transcranial Magnetic Stimulation in Clinical Practice and Research." *Clinical Neurophysiology* 120 (12): 2008–2039.

At the time of this writing, the definitive, authoritative compilation of safety and ethical information relating to the use of TMS in cognitive neuroscience and clinical research. It replaces an analogous report that had been published in 1998, and so one can imagine that the next update will appear a year or so preceding 2020.

Walsh, Vincent, and Alvaro Pascual-Leone. 2003. *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind*. Cambridge, MA: MIT Press.

A thorough, readable overview of the physics and physiology underlying TMS, and the cognitive neuroscience applications of this technique.

CHAPTER 4

AUDITION AND SOMATOSENSATION

KEY THEMES

- Auditory perception depends on exquisitely delicate and precise mechanical transmission in the inner ear, including the cochlea.
- The physical properties of the basilar membrane within the cochlea result in a spectral decomposition of acoustic energy, which is read out as a place code by the auditory nerve.
- The multiple synaptic connections of the auditory brainstem produce the stereotyped early- and middle-latency components of the auditory evoked potential (AEP).
- Deriving event-related potentials (ERP) from the EEG, such as the AEP, is a powerful way to measure neural events that are difficult to isolate in the “noisy” EEG.
- Primary auditory cortex (A1), in the superior temporal gyrus, contains a 1-D tonotopic map.
- Many different types of receptors are responsible for the detection of mechanical vs. thermal vs. painful sensations on the skin.
- S1 contains somatotopic maps, distributed across four Brodmann areas in postcentral gyrus of the primate.
- Optogenetic manipulation of S1, combined with electrophysiology, has provided insight into the functions of, and mechanisms underlying, gamma-band oscillations in the LFP.
- Discoveries about somatosensory plasticity in the rodent and nonhuman primate have provided insight into the neural bases of learning, as well as the neurological disorders of phantom limb and phantom pain.

CONTENTS

KEY THEMES

APOLOGIA

AUDITION

Auditory sensation

What is it that we hear?

Mechanical transmission and neural transduction

Spectral decomposition performed by the basilar membrane

Auditory perception

The auditory evoked response

The organization of auditory cortex

Top-down control of auditory transduction

Age-related hearing loss

Adieu to audition

SOMATOSENSATION

Transduction of mechanical and thermal energy, and of pain

Somatotopy

Somatotopy in the rodent: barrel cortex

Somatosensory plasticity

Use-dependent plasticity

The brain mimics the body

Mechanisms of somatosensory plasticity

Phantom limbs and phantom pain

A hypothesized mechanism

Helpful or harmful?

Proprioception

Adieu to sensation

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

APOLOGIA

If sensory systems were people whose feelings could be hurt, I would feel apologetic, and guilty, toward audition and somatosensation, and even more so toward olfaction and gustation, for the short shrift that they receive in this book. In this author's defense, however, I think it can fairly be said that such "visuocentrism" is characteristic of how most cognitive neuroscience texts are written and courses taught, with the exception of those that focus on a particular sensory modality or a domain of cognition, such as language, that is heavily dependent on a sense other than vision. (And, of course, within language, reading is a subdomain that is highly reliant on vision ...) The hope for this chapter, as, indeed, is true for all chapters in this text, is that the necessarily selective overview that is offered here will prompt some to want to study these problems in more detail. Where possible, the intent will be to emphasize principles and themes that generalize to other areas of cognitive neuroscience.

AUDITION

Audition has been studied from many perspectives, including speech perception, sound localization, and plasticity. This chapter will focus on the mechanical and signal-processing elements of sound transduction, an overview of auditory system anatomy, and the cortical representation of sound.

Auditory sensation

What is it that we hear?

Just as vision results from the transduction of energy from a relatively narrow band of the electromagnetic spectrum (what we call "visible light"), audition is the transduction of energy in the form of changes in air pressure. Many are familiar with the question *If a tree falls in a forest, but there are no humans within one hundred miles, does it make a sound?* Auditory science tells us unequivocally that the answer is "No." Consider this: if a human were on the surface of an asteroid floating in the vacuum of space, and it were possible to wear a spacesuit that left the ears uncovered, the most violent

of collisions (falling tree, exploding supernova, whatever) would not make a sound. This is because the auditory system works on the principle of detecting changes in energy in the medium making contact with the ear. In everyday life, this corresponds to the movement of molecules of the gases that make up the air that's all around us; or, when underwater, to collisions between and movement of molecules of water (plus whatever else may be dissolved in it). We will refer to this as **acoustic energy**. (Publicity for the sci-fi movie *Alien* had it right: "In space, no one can hear you scream.")

Mechanical transmission and neural transduction

From a mechanical standpoint, the steps leading to transduction of this energy are much more complicated than are those involved in the transduction of light energy. These are illustrated in *Figure 4.1*, but, in summary, a membrane in the outer ear (the tympanic membrane) vibrates in response to vibrations in the air, and its vibration is transferred through a series of tiny bones (the ossicles) to another membrane, the oval window, whose vibration produces changes of pressure in the fluid-filled cochlea. From the outside, the cochlea resembles a tube that has been rolled into

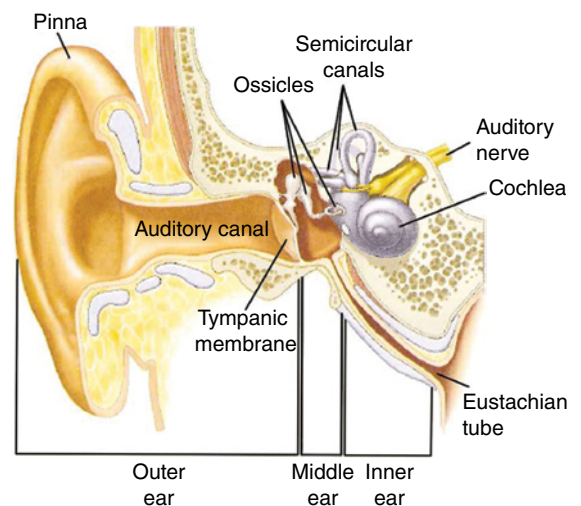


FIGURE 4.1 Major structures of the outer, middle, and inner ear. Source: Yantis, 2014. Reproduced with permission of Worth Publishers.

a spiral. (Indeed, *cochlea* derives from the Greek for snail.) There are, in fact, three long and narrow chambers rolled up within the cochlea, and we'll focus on the (long and narrow) membrane that separates two of these chambers – the **basilar membrane**. The basilar membrane vibrates in response to the pressure changes caused by the “pushing” and “pulling” of the ossicles on the oval window. Note that, up to this point, what we've summarized is a series of mechanical *transfers* of energy, from vibrations in the atmosphere, through several intermediate steps, to vibrations of the basilar membrane. *Transduction* hasn't yet occurred.

Transduction of the mechanical, vibrational energy of the basilar membrane into a neural signal occurs in **inner hair cells** that are attached all along the long-axis of the basilar membrane (*Tip 4.1*). At a simplified level of description, the base of each inner hair cell is secured to the basilar membrane. At the other end of the cells are tufts of hair-like cilia extending into the fluid of the cochlea. Movement of the basilar membrane causes the cilia of inner hair cells to bend, “like tiny clumps of seaweed swept back and forth by invisible underwater currents” (Sekuler and Blake, 1985, p. 311). This back-and-forth movement produces mechanical shearing forces in the membrane of the hair cell, which, in turn, opens or closes mechanically gated ion channels. The resting potential of inner hair cells is approximately -60 mV , and the opening of these channels results in the inflow of K^+ and Ca^{2+} ions, the consequent depolarization of the cells, and the consequent release of neurotransmitter (probably glutamate) at synapses linking the base of the inner hair cell to **ganglion** cells of the cochlear nucleus (*Tip 4.2*). These ganglion cells are the first neurons of the auditory system, analogous to

Tip 4.1

Also a part of the mechanics of the cochlea are outer hair cells, which play a role of “tuning” the vibration of the basilar membrane so as to increase the signal-to-noise ratio of the signals transmitted by inner hair cells to the neurons of the cochlear nucleus.

Tip 4.2

The cochlea is very difficult to study and, probably due to differences in the geometry of the structure in different species, we know more about the detailed workings of the cochlea – including which neurotransmitters are released by hair cells – in the turtle than in mammals.

the retinal ganglion cells of the visual system, and they carry electrical signals about auditory information into the central nervous system. Together, the fibers of these cochlear ganglion cells are known as the **auditory nerve**. (We'll return to **outer hair cells** further along in this chapter.)

Two important properties of sound are its volume and pitch. Volume is encoded in this system in a fairly straightforward way: increasing the volume of a sound results in greater displacement of the oval window (i.e., “bigger” vibrations) and, consequently, in greater deflection of the basilar membrane, greater shearing forces on the cilia, greater depolarization of the inner hair cells, and, finally, the release of more glutamate onto cochlear ganglion cells. The end result is that the volume of a sound is encoded by the rate at which neurons of the cochlear nucleus fire (i.e., a **rate code**). To understand pitch, we have to understand another fundamental mechanical property of the basilar membrane, which is the exquisite specificity with which different regions vibrate as a function of frequency.

Spectral decomposition performed by the basilar membrane

Having drawn the analogy between cochlear ganglion cells and retinal ganglion cells, it is important to note differences in how they encode information, differences that arise from fundamental differences in the physical properties of photons vs. acoustic energy. For our purposes, photons travel in a straight line, a property that our visual system exploits to detect objects regardless of where they appear in the visual field. Thus, for example, a visible rustling of leaves in a tree will produce a pattern of changing light energy on any small patch of the retina, regardless of whether one's center of gaze is directed at the tree, to the right of it, above it, or below it. Thus, retinal ganglion cells

carry signals that, when pooled, create a 2-D map of space. Acoustic energy, in contrast, expands out from its source in all directions. (A commonly used metaphor is the pattern of expanding, concentric ripples that result from dropping a pebble into a previously placid pond – just keep in mind that acoustic “ripples” expand in 3-D, in concentric spheres.) Therefore, acoustic energy would not project in an orderly way against a 2-D receptor sheet in the way that visible light energy projects onto the retina. Consequently, the inner ear has evolved to discriminate differences in auditory stimuli along just one dimension (other than volume), and that’s vibrational frequency. The physical property of vibrational frequency corresponds to the auditory percept of pitch, such that low-frequency vibrations (as produced, e.g., by a tuba) are perceived as having low pitch, and high-frequency vibrations (as produced, e.g., by a flute) are perceived as having high pitch.

Although the cochlea is a spiral structure, it’s useful to think of it as a (“unrolled”) straight tube. The end of the tube where the oval window is located is referred to as the base. The tube is widest at this end, and it progressively narrows until it terminates at the apex. Interestingly, however, the basilar membrane within this tube is thick and narrow at the base, and becomes progressively thinner and wider as one moves toward the apex. The consequence of this configuration is that different regions of the basilar membrane will vibrate preferentially to acoustic vibrations of different frequencies. In effect, the basilar membrane is tuned like a piano, with the regions near the base vibrating preferentially to the highest-frequency sounds that we can hear, and regions near the apex vibrating preferentially to the lowest-frequency sounds that we can hear (*Figure 4.2*). This results in a principle for representing sounds of different frequencies with what is referred to as a **place code**, reflecting the fact that auditory nerve fibers arising from different *places* along the basilar membrane preferentially carry information about one relatively narrow range of frequencies.

If you understand the ideas that have been described in the two previous paragraphs and accompanying figures, then you understand the concepts underlying the “spectral decomposition” referred to in this section’s title. What’s left is for us to do is to match those concepts

Tip 4.3

Note that although spectra can cover a large range of frequencies, the nervous system can typically only perceive a relatively narrow range within that spectrum. Thus, there are many examples of acoustic energy in the environment whose frequencies are either too high or too low for humans to detect.

onto this jargony phrase. The frequency at which a signal oscillates, such as the physical vibrations that produce acoustic energy, can be characterized as falling along a continuum, or a *spectrum*, that ranges from low frequencies to high frequencies (*Tip 4.3*). A sound source that vibrates at just one frequency is referred to as a pure tone. (This would be represented mathematically by a sine wave; see *Methodology Box 4.1*.) As we’ve established, a pure tone with a relatively low pitch would produce a traveling wave whose largest peak occurs at a single location near the apex end of the basilar membrane, whereas a pure tone with a relatively high pitch would produce a peak at a location near the base end. A new piece of information to add to this discussion is that if we were to play both of these tones simultaneously, the resultant “combined” traveling wave would produce two peaks, one at each of these two locations along the basilar membrane. Indeed, most sounds in nature are made up of a combination of many, many vibrational sources, each vibrating at a different frequency, as well as with a different amplitude. Consider for instance the human voice (a sound source that we won’t consider in detail until *Chapter 18*). Speech is produced by forcing air across many flaps of tissue (the vocal cords), each of which vibrates at a different frequency; vibrations of the uvula, tongue, and lips also contribute. This *mélange* of vibrations at multiple frequencies is conveyed by a single complicated pattern of pushes and pulls by the oval window. This is a point that is important to emphasize: because there is only one oval window, there is no sense in which many *distinct* vibrational signals are being delivered to the cochlear nucleus from different sources. Rather, what the cochlear nucleus receives is a *single* signal in which many different frequencies are embedded. Although this amounts to a much more complicated version of

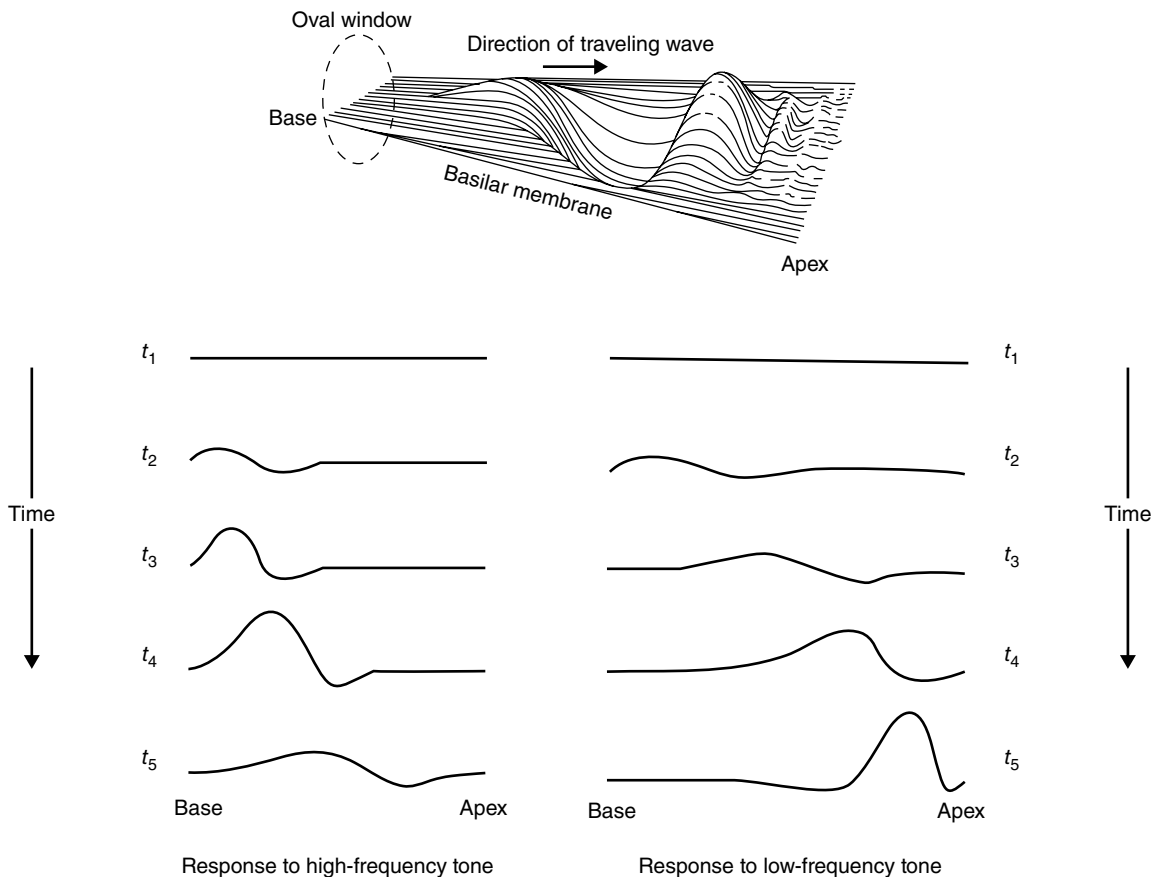


FIGURE 4.2 Vibrations of the oval window create **traveling waves** along the basilar membrane. (An example of a traveling wave is what happens if you tie a rope to a pole, walk away with the other end in your hand until the rope no longer touches the ground, then snap your wrist as though you are cracking a whip.) The top panel illustrates a moment in time as the basilar membrane is responding to a sound. The bottom panel represents the response of the basilar membrane at five points in time (t_1 through t_5) after the playing of a high-frequency tone (left-hand column) and a low-frequency tone (right-hand column). This illustrates the fact that, unlike the “whipped rope” example, traveling waves in the cochlea are amplified or attenuated as they travel the length of the basilar membrane such that the wave corresponding to any one frequency of vibration of the oval window will only peak at one location along the basilar membrane (see *Tip 4.1*). Further along in this chapter we’ll see that it’s not only the physical structure of the cochlea, but also the activity of the outer hair cells that produce this distinctive property. Source: Sekuler and Blake, 1985. Reproduced with permission of Alfred A. Knopf, Inc.

our scenario of playing two pure tones simultaneously, the basilar membrane responds to it in the same way – the multiply-embedded traveling wave contains multiple peaks along the length of the basilar membrane, one corresponding to each vibrational frequency in the sound. It is in this sense that the basilar membrane can be thought of as decomposing the sound signal: the

signal comes in as a single “complex waveform” in which many different vibrational frequencies are embedded, and it “leaves” the cochlea (via the auditory nerve) having been separated (“decomposed,” in the parlance of signal processing) into its constituent frequencies. (Because each frequency can be represented as occurring at some point along a spectrum, the

METHODOLOGY BOX

4.1 Spectral analysis of time-varying signals

In the early 1800s, the French mathematician Jean Baptiste Joseph Fourier (1768–1830) proposed that any fluctuating waveform, no matter how complicated, can be represented as a finite collection of sines and cosines. The mathematical operation of transforming a fluctuating time series into a collection of sinusoids has come to be known as performing a **Fourier transform** (see *Tip 4.4*). The resultant collection of sinusoids can be represented in a **power spectrum**, as illustrated in Figure 4.3.D, and it is thus also known, in jargon, as a *spectral transform*. It would be difficult to overstate the impact that Fourier's theorem has had on the natural and physical sciences, and on many branches of engineering. From a practical perspective, sinusoids and power spectra can be easier to work with analytically than are complicated signals that fluctuate over time. From a conceptual standpoint, the idea that many systems can be thought of as having cyclical (or oscillatory) properties has provided many new insights for how these systems might work. We have already discussed one example of this in the neurosciences: viewing fluctuations in the EEG as the waxing and waning of oscillatory magnitude in different frequency bands gives rise to the idea that the visual system can be in different states, and that the momentary state of the system has important consequences for how it is functioning, as well as for how it may respond to, say, a change in the environment.

As we proceed through this book, it will be helpful to be familiar with some concepts, conventions, and terminology related to the spectral analysis of time-varying signals. (For starters, a more efficient way to say “the spectral analysis of time-varying signals” is “signal processing,” a term that originated in electrical engineering.) *Figure 4.3* illustrates many of these. *Figure 4.3.A* illustrates EEG signal from electrode P1 during the delay period of a short-term memory task (something we'll take up in detail in *Chapter 13*). Note that this reflects the average of several trials. *Figure 4.3.B* illustrates how the waveform in *A* can be decomposed into a series of sinusoids of different frequencies. Note that this is only a subset of the frequencies required to completely represent the waveform in *A*; the full complement of sinusoids is represented in the power spectrum (*Figure 4.3.D*). Returning to *Figure 4.3.B*, note that, in addition to frequency, the sinusoids illustrated here differ from one another in two other ways. One is that each differs in terms of its magnitude: the height of its peaks and the depth of its troughs. In the parlance of signal processing, the magnitude of a sinusoid is referred to as its **power** (*Tip 4.5*). The second is that each of the sinusoidal waves is at a different phase of its cycle at time 0. The power spectrum illustrated in *Figure 4.3.D* integrates across the 3000 ms of the delay period. Another way to represent these data is to, in effect, compute the Fourier transform at many narrower slices in time, and thereby illustrate the evolution of the spectral representation over time, in a time-frequency **spectrogram** (*Figure 4.3.E*). One practical application of these principles is the design of cochlear implants, as considered in *Real-World Applications 4.1*.

Methodology Box 4.1 Continued

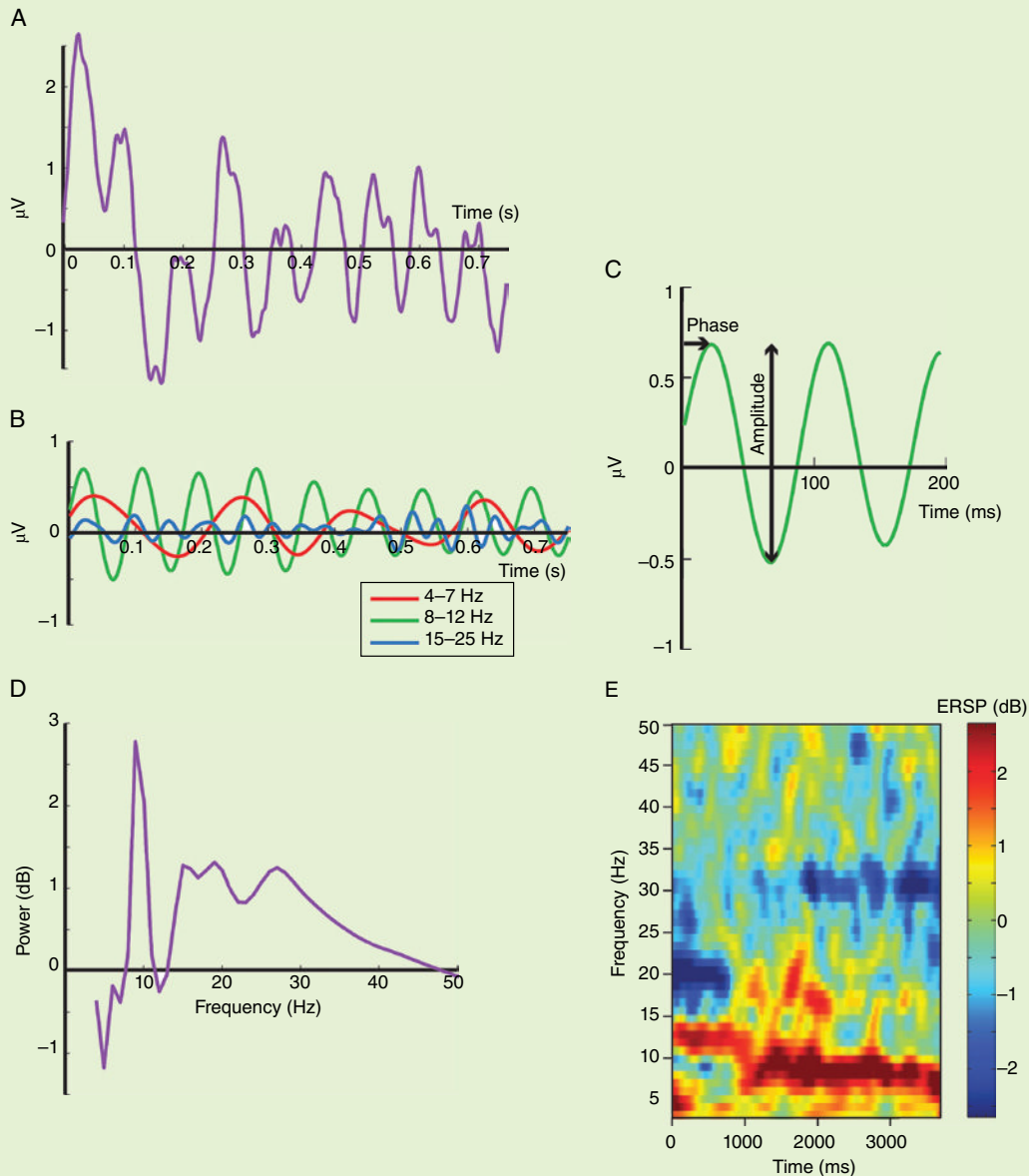


FIGURE 4.3 Spectral decomposition of EEG data. **A.** A 700 ms snippet of EEG data. **B** shows the “spectral decomposition” of the data from A into sinusoids from three frequency bands: theta (4–7 Hz); alpha (8–12 Hz); and beta (15–25 Hz). Note that there are also frequency components in A that are lower and higher than the three frequency bands shown here. **C.** Illustration of the properties of phase and amplitude of a sine wave. **D.** The power spectrum from the 3 second delay period of the short-term memory task as illustrated in E. Note that this power spectrum reflects the task-related change in oscillatory power vs. baseline; thus, the fact that it dips below 0 at low frequencies means that there is more low-frequency power in the EEG during baseline than during task. **E.** The spectrogram from the delay period of a spatial short-term memory task, with power at each frequency bin at each point in time computed as an “event-related spectral perturbation” (ERSP), and expressed in units of dB. (Details of the task that produced these data are considered in *Chapter 8* and *Chapter 13*.) Source: Figure courtesy of Bornali Kundu, derived from data from Johnson, Kundu, Casali, and Postle, 2012.

Tip 4.4

There are different ways, computationally, to perform a spectral transform. One that is computationally efficient, and therefore often used in science and engineering, is known as the fast Fourier transform (FFT).

Tip 4.5

As you might imagine, this section is skipping over semesters' worth of mathematical details, one of which is that whereas the band pass filtered sinusoidal waves in *Figure 4.3.B* are measured in μV , power is expressed in units of $\mu\text{V}^2/\text{Hz}$ or decibels (dB).

decomposition of a complex waveform into constituent frequencies can be called a “spectral decomposition.”) In effect, the cochlea has performed an operation on the sound signal that is equivalent to the mathematical operation known as the Fourier transform (*Methodology Box 4.1*). Visit *Web Link 4.1* for a remarkable visual simulation of how the basilar membrane responds to sounds, including a snippet from Bach's Toccata and Fugue in D Minor.

Auditory perception

The auditory evoked response

As *Figure 4.4* illustrates, the auditory signal is processed in many subcortical structures, collectively referred to as the **brainstem auditory nuclei**, before arriving at primary auditory cortex, A1. Synaptic activity associated with these many nodes produce the many deflections (the “peaks”) in the early-latency components of the auditory evoked potential (AEP; *Figure 4.5*). It is in these brainstem nuclei that many of the computations are performed that allow us, among other things, to localize where a sound in the environment is coming from (*Tip 4.6*). Importantly, at each of these nodes, the topography of the “place code” for frequency that was established by the properties of the basilar membrane is preserved. The remarkable speed at which all of this

Tip 4.6

Sound localization presents a fascinating engineering problem, one that is solved very elegantly by the circuitry of the auditory brain stem. As a thought problem, how could you design a system to localize the source of a sound if you only had two detectors (i.e., two ears), and so the only two pieces of relevant information available to you would be: (a) the relative intensity and (b) the relative time of detection of the sound by each of the two detectors? The former is referred to as the **interaural intensity difference (IID)** and the latter as the **interaural timing difference (ITD)**.

processing takes place can be appreciated from inspection of the AEP (*Figure 4.5*), each peak of which is associated with a synaptic event (and, therefore, a stage of processing) along the auditory pathway.

To foreshadow some questions that we'll consider in future chapters, let's note that sensory evoked responses (the AEP and its visual and somatosensory analogues) are stereotyped, stable responses that are effectively identical in all neurologically healthy individuals. This makes them very effective “tools” for studying the neural bases of many different behavioral and physiological states. To study attention, for example, how does the auditory evoked response differ on trials when one is attending to a concurrent stream of visual stimuli vs. trials when one is attending to the auditory stimuli? (For example, does inspection of *Figure 4.6* suggest that some steps in the neural processing of sound are vs. are not influenced by attention?) To study different stages of sleep, how does it differ when tones are played when one is awake (and quietly resting), vs. in stage 1 (i.e., “light”) sleep, vs. stage 2 (i.e., “heavy”) sleep, vs. rapid-eye-movement (REM) sleep?

The organization of auditory cortex

In the AEP (*Figure 4.5*), the N1 and subsequent components are the first that unambiguously correspond to cortical processing. Thus, despite all the circuitry of the auditory brainstem, auditory information gets from ear to cortex in considerably less

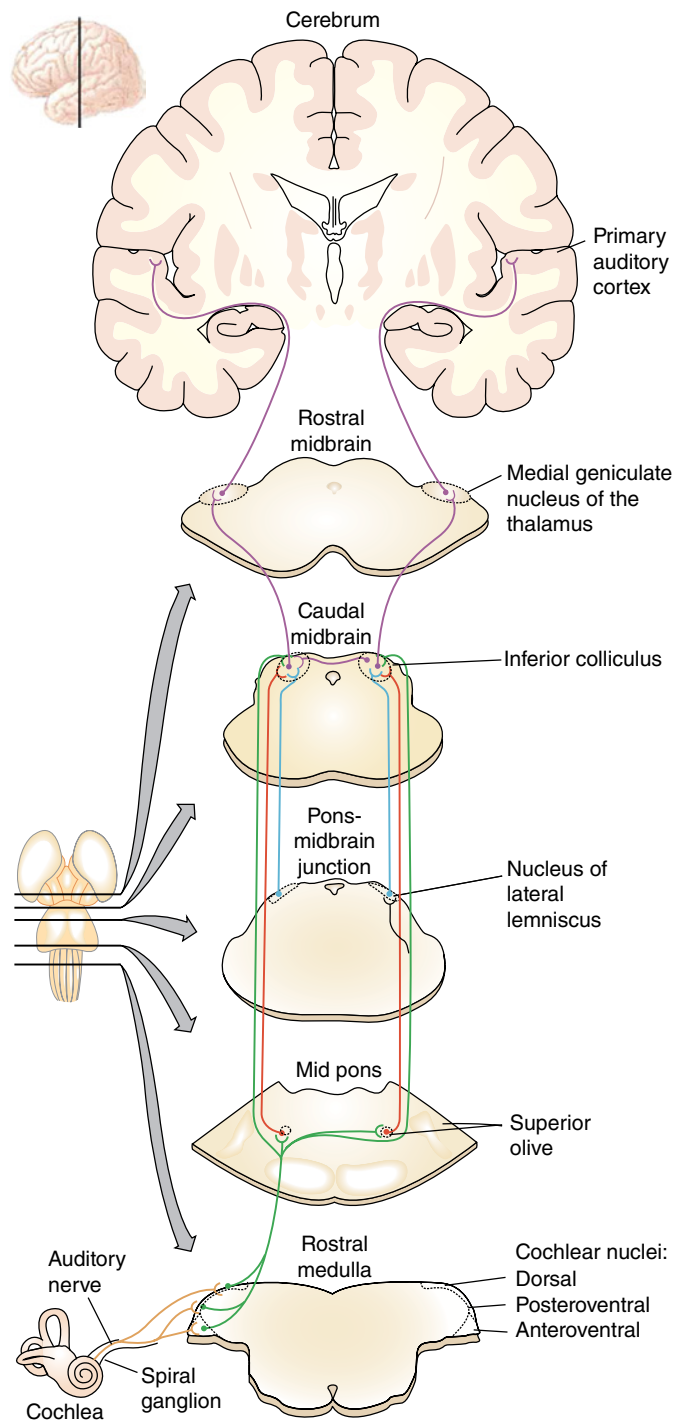


FIGURE 4.4 The auditory pathways from cochlea to cortex. On the left-hand side of the figure are three 3-D renderings of anatomical structures: the bottom is the cochlea; in the middle is a view from behind of the brainstem, midbrain, and thalamus (with cerebellum removed), and horizontal lines showing the locations of the cross-sectional slices illustrated on the right; and at the top is a view of the cerebrum, with a line indicating the level of the coronal slice. Note that there are multiple pathways, with synapses at each of the five levels of brainstem and midbrain that are illustrated (although not every pathway synapses at each level). Pathways originating from only the left ear are shown. Source: Wolfe, Kleunder, Levi, Bartoshuk, Herz, Klatzky, Lederman, and Merfeld, 2012. Reproduced with permission of Sinauer Associates, Inc.

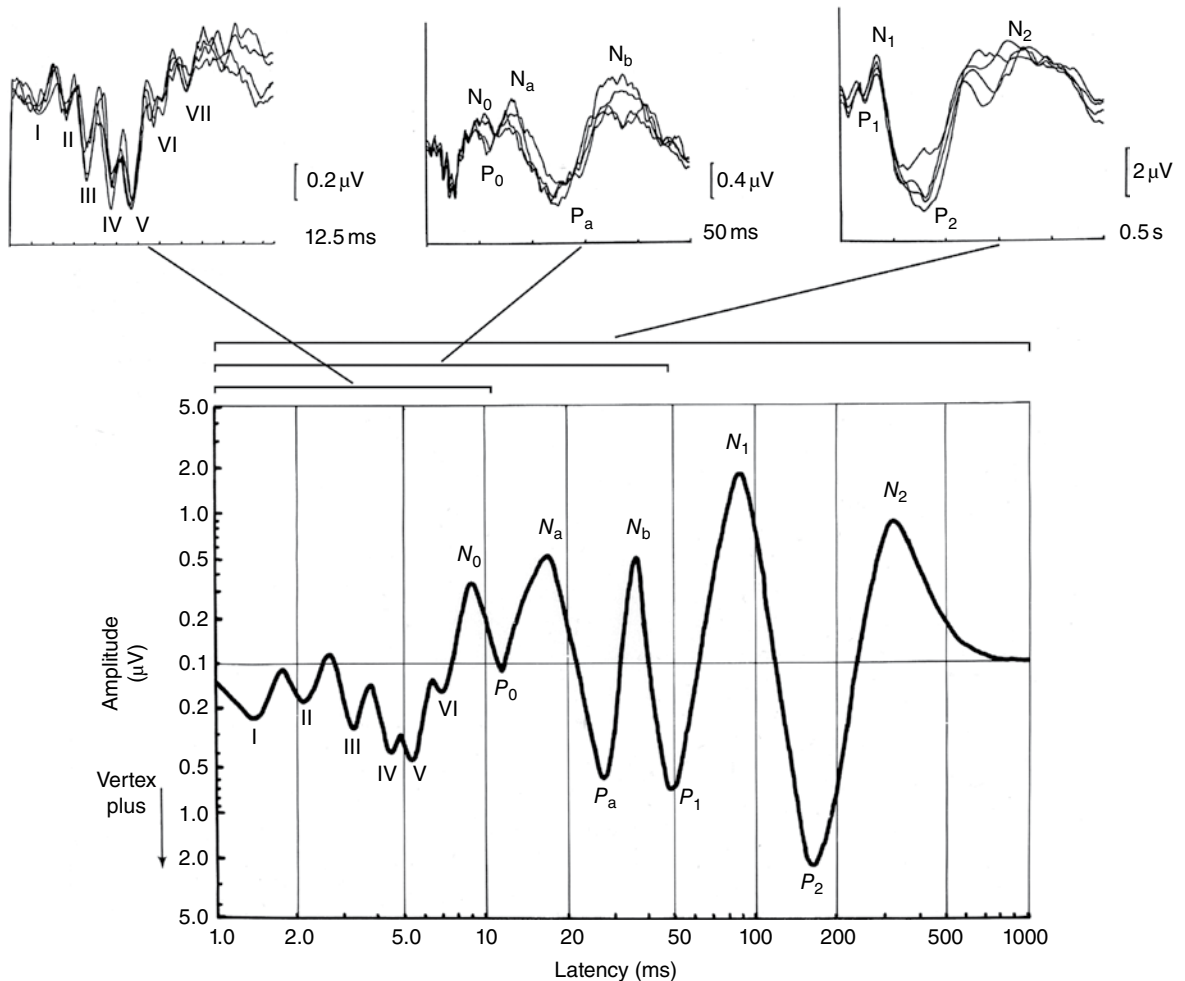


FIGURE 4.5 The auditory evoked potential (AEP) to a 60 dB_{SPL} click of 50 μs duration. Each of the panels in the top row shows four traces, each an ERP computed by averaging the EEG from 1024 trials, and each acquired from the same subject, but in four different recording sessions. (This was done to demonstrate stability of this response within an individual). Although each of these waveforms is timelocked to the click, each differs with regard to the length of time following the click over which the EEG traces were averaged. Thus, the panel on the left illustrates the first 12.5 ms post-click; the middle panel includes these initial 12.5 ms, but also the ensuing 37.5 ms (i.e., a time span of 50 ms); the panel on the right a time span of 500 ms. Note that the scale on the vertical axis (i.e., the amplitude of the ERP) changes by an order of magnitude from the 12.5 ms time window to the 500 ms time window. These three panels correspond to what are known as the *early*, *middle*, and *long* latency components of the AEP, with the early latency components known to correspond to processing in the auditory brainstem structures illustrated in Figure 4.4. The bottom panel illustrates the AEP (windowed at 1000 ms) averaged across eight individuals, and accommodates the fact that these three groups of components play out on different scales of time and magnitude by plotting the “log/log” transform of the AEP (that is, both the temporal and amplitude values are plotted on logarithmic scales). Source: Adapted from Picton, Hillyard, Krausz, and Galambos, 1974. Reproduced with permission of Elsevier.

REAL-WORLD APPLICATIONS

4.1 External hearing aids and cochlear implants

Although there are many classes of hearing impairment with varying **etiologies**, here we will only distinguish between the broad categories of hearing loss, one due to impaired mechanical transmission of sound, and one due to impaired transduction. The former accrues from the “wear and tear” of daily life (and varies considerably depending on one’s exposure to loud sounds), and is an inevitable consequence of aging. The most common consequence of partial damage to the machinery responsible for the mechanical transmission of sound, as well for the mechanical working of the inner ear, is to raise the threshold at which sounds can be detected. Thus, the most common treatment for this kind of hearing loss is amplification. The traditional hearing aid, which sits in the auditory canal (i.e., between the ear and the tympanic membrane), works as a filter that detects the same vibrations as does the tympanic membrane, but selectively passes and amplifies only those frequencies that the user wants to emphasize, typically those that carry human speech (250–4000 Hz). Importantly, the amplification is dynamic, such that a sound of 20 dB is amplified much more than a sound of 90 dB, for the simple reason that intensity thresholds at which the inner ear can sustain damage and/or at which an individual can experience pain are no different for someone experiencing transmission-related hearing loss than for someone who is not.

The second class of hearing loss is known as sensorineural. Most often this is due to physical damage to (or abnormality of) the hair cell–cochlear nucleus interface, loss of hair cells (as can result from certain antibiotics or cancer drugs), or damage to the auditory nerve. An amplification-based hearing aid is useless for sensorineural hearing loss. (By analogy, if an individual is born with or has sustained damage to the retina, simply making a computer monitor brighter will have no effect on the individual’s inability to see it.) If, however, the damage primarily involves the hair cells – a common culprit with sensorineural hearing loss – cochlear prosthetics offer one treatment option. A “cochlear implant,” as it is commonly known, is effectively a flexible stimulating electrode with many contacts along its shaft. It is surgically fed into the cochlea so that it coils in the same way as does the basilar membrane, and so that the electrode contacts make physical contact with fibers from the cochlear nucleus. A microphone that sits behind the ear transmits information about the acoustic signal to a surgically implanted microprocessor that, in effect, seeks to emulate the place code of the cochlea by delivering patterns of microstimulation (via the implanted electrode) that mimic how the basilar membrane would respond to that sound.

Although different patients experience differing degrees of benefit from cochlear implants, one general principle is that the procedure is most successful when performed in very young children (*Tip 4.7*). This is presumably because the infant brain is more plastic than that of even primary-school-aged children, and so the circuitry of the infant auditory cortex is better able to interpret the “unnatural” signal generated by the cochlear implant (*Web Clip 4.1*; but wait until you’ve read this chapter’s section on experience-driven plasticity.)

than 100 ms! The primary auditory cortex, A1, is located in posterior superior temporal gyrus, bilaterally, on the ventral wall of the Sylvian fissure (*Figure 4.7*). In a manner that echoes the functional organization of primary visual cortex, the place code that originates in the cochlea is preserved all

the way to A1. But although the analogy to retinotopy might make one expect that A1’s functional organization would be called “basilotopy” or “cochleotopy,” the standard term is **tonotopy**, a reference to the fact that it is organized in terms of the frequency of the tone(s) being processed. That is, as

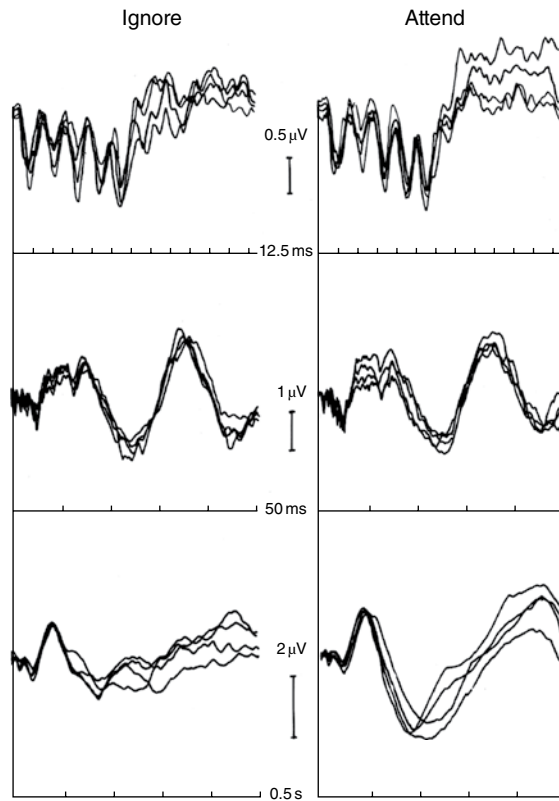


FIGURE 4.6 How attention to sound does, and doesn't, influence the AEP. Clicks ($60 \text{ dB}_{\text{SPL}}$) were delivered once per second, and once every 5–30 seconds the click was slightly softer. In the IGNORE condition the subject was instructed to read a book and pay no attention to the clicks. In the ATTEND condition, s/he was instructed to detect each softer click and report the number detected at the end of the session. This figure illustrates the AEP to the soft click using the same three time windows as *Figure 4.5*, but here they are arranged vertically to facilitate comparison between the IGNORE and ATTEND conditions. An effect of attention was only seen reliably on the N1 and P2 components of the AEP (bottom row). The figure illustrates two important points, one relating to the effect of attention in the ATTEND condition, the second to the absence of detectable effects of reading in the IGNORE condition. In the ATTEND condition, because there was not an external event occurring roughly 100 ms after the click, the effects on the N1 and P2 ERP components are assumed to result from a *neural event* that was not directly evoked by an external stimulus. That is, they reflect an endogenous influence on the ERP. (We'll consider the nature of this neural event, and other attention-related ERP components, in *Chapter 6*.) The ERP from the IGNORE condition, in contrast, illustrates how trial averaging can remove components from the EEG that derive from events that one is not interested in. Note how the waveforms from the two conditions look nearly identical (apart from N1 and P2), even though only one includes reading. Reading involves regular eye movements, which produce enormous artifacts in the EEG, as well as neural processing associated with visual perception (*Chapters 3 and 5*) and analyzing the grammar and the meaning of sentences (*Chapter 18*). However, because none of these reading-related events was timelocked to the clicks, their influence on the EEG was averaged out of these ERPs. Source: Picton, Hillyard, Krausz, and Galambos, 1974. Reproduced with permission of Elsevier.

Tip 4.7

There is controversy in the deaf community about the ethics of implanting cochlear implants in the very young. Deaf culture is very rich (see *Further Reading* by Padden and Humphries, 2005), as are sign languages, and the concern is that a sociocultural outcome of the surgery may be to isolate the child both from deaf culture (because the choice makes it less likely that the child and her/his family will learn sign language) and from mainstream culture (because even the most successful cochlear implants leave the patient with marked hearing impairments).

one moves along the long axis of the surface of A1, one finds that neurons respond to progressively higher (or lower) frequencies. (Note that A1 can be considered to have a “1-D” representation of sound, in that preferred frequency does not vary appreciably along the mediolateral axis of A1, but only along its rostrocaudal axis.) Also analogous to the visual system is the fact that neurons in A1 project to anatomically adjacent areas – referred to as “belt” and “parabelt” cortex – for subsequent processing. Unlike the visual system, however, auditory cortex has systematic differences between the two hemispheres, both in terms of its anatomy and the kind of information that it processes. To provide a hint about one of the important factors behind these differences, the question of lateralization will be addressed in some detail when we consider the neural bases of speech perception, in *Chapter 18*.

Top-down control of auditory transduction

Before leaving the auditory system, we revisit the principle of feedback in sensory processing. Not unlike the optic radiations, which carry more axons traveling from cortex to LGN than from LGN to cortex, the auditory nerve carries more fibers projecting to the cochlea than projecting from it. Of course, an important difference between these two systems is that the cochlea is the site of transduction of acoustic energy into a neural code, more analogous to the retina than to the LGN. (The structural

analogue of the LGN in the auditory system is the medial geniculate nucleus of the thalamus (MGN; *Figure 4.4*.) Indeed, whereas the role of top-down projections to the LGN is to implement neural filtering of the afferent visual signal, in the auditory nerve the fibers projecting into the cochlea are directly influencing the very mechanism of transduction. They do so by synapsing on the outer hair cells, thereby controlling the behavior of these cells. Outer hair cells, although also attached to the basilar membrane at one end and sprouting cilia at the other, differ from inner hair cells in that their cilia attach to another part of the cochlea. A result of the coordinated activity of many outer hair cells, therefore, can be to sharpen the movement of the basilar membrane by dampening its vibrations on both sides of the peak of the traveling wave. In effect, the outer hair cells implement a mechanical center-surround inhibition of the movement of the basilar membrane. An important focus of ongoing research in auditory perception is determining the source and the nature of the control signals to the outer hair cells.

Age-related hearing loss

One subdiscipline of cognitive neuroscience that won't get much direct attention in this book is cognitive aging. Nonetheless, as *Real-World Applications 4.2* illustrates, it is one of the areas where our field interfaces directly with a question of broad societal import. This content is supplemented with perspectives on the neurophysiology of aging in *Web Clip 9.2* and *Web Clip 10.2.B*.

Adieu to audition

This concludes our whirlwind introduction to auditory perception. Some of the principles that we've covered here will be relevant to this chapter's next section, on somatosensation, and others to many topics that will be covered further along in this book. Additionally, of course, the auditory system will be front-and-center in *Chapter 18*'s focus on the cognitive neuroscience of language. There, among other topics, we will consider how speech is processed relative to nonspeech-related sounds. But now, onto the sense of touch.

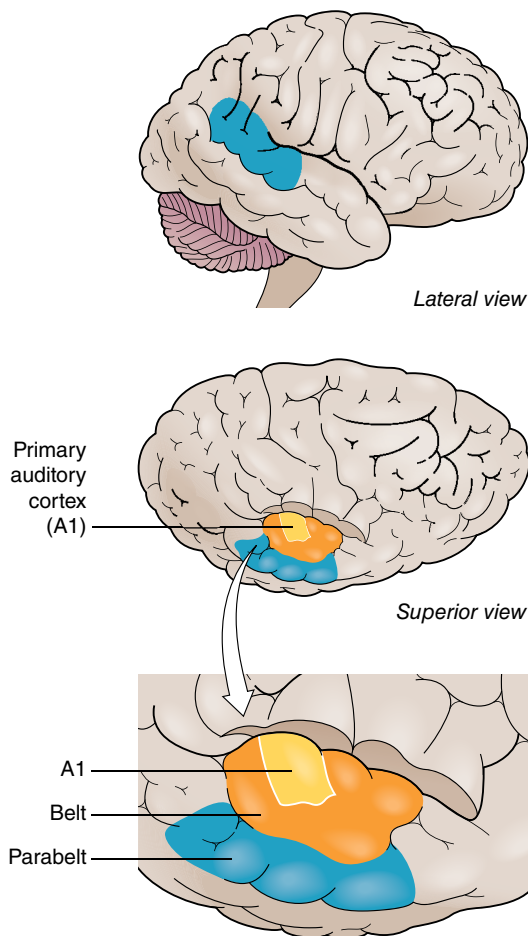


FIGURE 4.7 A1, belt, and parabelt cortex of the human auditory system (shown for the right hemisphere). Because these lie on the ventral bank of the Sylvian fissure, the overlying parietal cortex has been “cut away” to expose them in the superior view.

SOMATOSENSATION

Transduction of mechanical and thermal energy, and of pain

The sensations that we “feel with our skin” result from the activation of one or several of the many types of somatosensory receptor cells embedded in the skin. These fall into three broad categories: **mechanoreceptors**, **thermoreceptors**, and pain receptors. Different mechanoreceptors are specialized for different mechanical

forces that can act upon the body: pressure, texture, flutter, vibration, and stretch. Distinct thermoreceptors respond to hot and to cold. Distinct pain receptors generate the sensation of pain due to extreme temperature, due to mechanical stress (e.g., pinching or cutting the skin), or due to chemical exposure. In addition to these functional properties, the different types of receptors summarized here also differ in terms of various physiological properties, such as size of receptive field, and speed of transduction and signal propagation. Each receptor is located at the end of a long process (technically, an axon) that projects past its cell body and into the spinal cord. The cell bodies of sensory receptor cells are clustered in ganglia that occur periodically along the long axis of the spinal cord. Because these ganglia are outside the spinal cord proper, these cells are classified as being a part of the peripheral nervous system (PNS; as opposed to the CNS). Indeed, unlike the prototypical CNS neuron, sensory receptor neurons do not generate action potentials at an axon hillock of the cell body. Rather, the action potential is generated in the membrane of the axon that is immediately adjacent to the receptor, and this action potential bypasses the cell body, propagating directly from the peripheral receptor into the spinal cord, and then rostrally along the cord to a synapse on one of two nuclei in the brainstem. Next, the ascending sensory signal is transmitted to the ventral posterior lateral (VPL) nucleus of the thalamus, and from here to the primary somatosensory cortex (S1; *Figure 4.8*). In the primate, S1 is located in parietal cortex immediately posterior to the central sulcus. Unlike either V1 or A1, S1 is divided across four distinct Brodmann areas, some of which support different functions (*Figure 4.9*).

Somatotopy

The principle of topographic organization of sensory cortex is perhaps most intuitively demonstrated in S1: adjacent parts of the body (“soma,” from the Greek) are, apart from a few necessary discontinuities, represented in adjacent regions of cortex. Although this organizational principle had been deduced from assessment of patients with focal brain lesions, it is most enduringly associated with the pioneering electrical stimulation studies of the Montreal-based neurosurgeon Wilder Penfield (1891–1976) and his colleagues. Because the brain itself is not **innervated**

REAL-WORLD APPLICATIONS

4.2 Aging and hearing loss

The week that I am writing this chapter, *The New York Times* has published a column from a woman of 65 years, Katherine Bouton, detailing her subjective sense of cognitive impairment accompanying hearing loss (Bouton, 2013). Some of what she describes is situational: at a crowded reception, she feels that the mental effort that she devotes to trying to hear her interlocutor precludes her from being able to recall the title of a book that she would have liked to have interjected into the conversation. Objectively, and understandably more troubling to Ms Bouton, there are also strong associations between the degree of hearing loss measured at one point in time (typically at around age 60), and subsequent declines on tests of cognitive function, as well as the likelihood of developing **dementia** (Alzheimer's disease or another type) in the ensuing decade (detailed in Lin et al., 2011, 2013). Logically, there are at least three models that could explain these associations: (1) the decline in hearing precedes, and causes, a subsequent decline in cognition; (2) the opposite is true; (3) hearing and cognition have no direct influence on each other, but a third factor that is common to both causes the parallel decline in both. By scenario #1, hearing loss might increase one's social isolation, and social isolation is a known risk factor for age-related cognitive decline and dementia. At the moment I can't think of a physiologically plausible story for scenario #2. Scenario #3, which could co-exist with either #1 or #2, could reflect any number of factors. One that has been hypothesized, but not yet rigorously tested, is cardiovascular disease. By this hypothesis, the integrity of the hair cells of the inner ear may be particularly vulnerable to systemic decline in the cardiovascular system (i.e., to abnormalities in blood flow that are experienced by the entire body). And so, like the proverbial canary in the coal mine, age-related hearing loss may be an early warning signal of gradually accruing damage to many other systems, including the brain.

It is certainly well established that age-related hearing loss can be exacerbated by damage to the intricate machinery of the inner ear that can accrue from hearing loud sounds earlier in life. To cite just one example, as Pete Townsend of *The Who* advances into his golden years, he has become an advocate of lowering the volume of headphones and earbuds, convinced that years of too-loud headphone use in the recording studio is responsible for his own severe hearing loss. (Perhaps, had he known at the time, he would have penned that notorious lyric as "I hope I die before I grow *deaf* ...".)

At the time of this writing, there are no definitive data on the factors discussed in this box. But what the column and an associated interview (Belluck, 2013) nicely capture is a layperson's perspective on (1) a personal concern about her neurological health; (2) how to navigate a literature in which there are many more active hypotheses than definitive answers; (3) the process of "doing science" (presenting preliminary results at conferences; submitting grant applications focused on a question that has arisen from one's previous research, etc.).

with somatosensory receptors, neurosurgery is often performed while the patient is awake, and can report what s/he is experiencing, and/or can be asked to perform actions that give the surgeon important moment-to-moment information about the status of the surgical intervention (*Tip 4.8*). Penfield was an epilepsy surgeon. His approach was guided by the

Tip 4.8

Of course, because the skin of the scalp is densely innervated with somatosensory receptors, topical anesthesia must be applied to the area(s) where skin and bone are retracted to expose the brain.

Ascending dorsal column–medial lemniscal pathway to primary sensory cortex

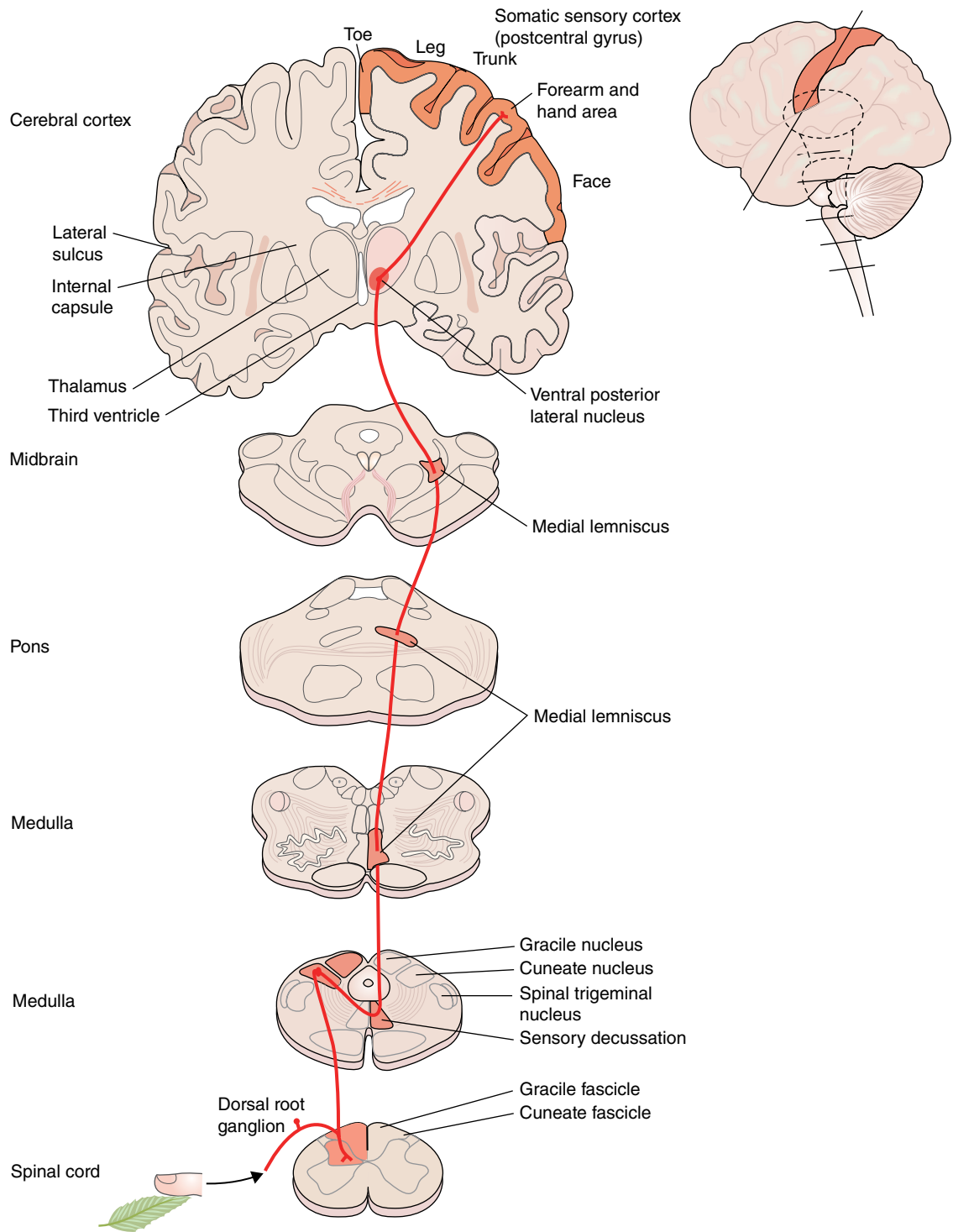
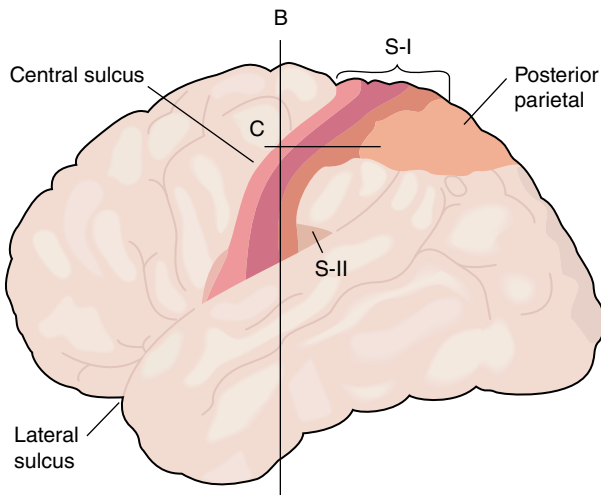
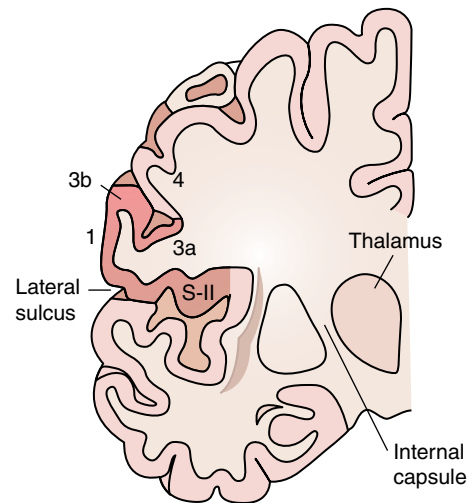


FIGURE 4.8 Somatosensory pathways from periphery to dorsal roots, ascending along dorsal spinal cord to ventral thalamus, and projecting into cortex. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

A
The somatosensory cortex



B
Coronal section



C

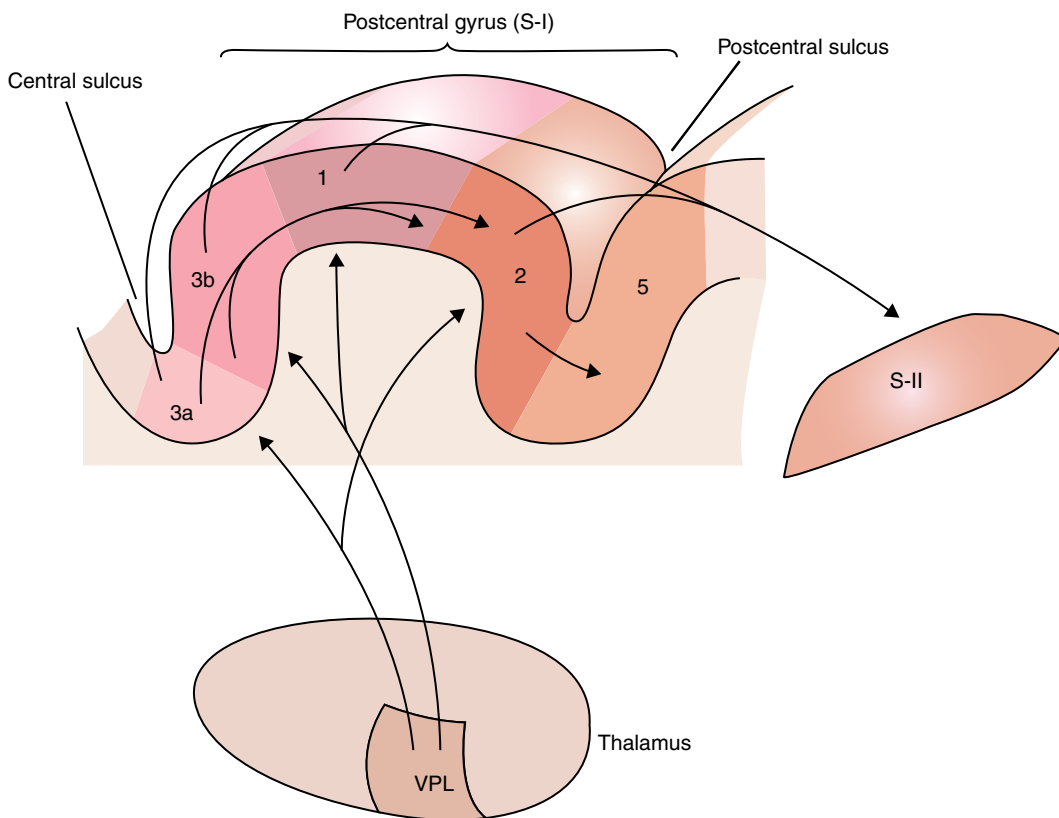


FIGURE 4.9 Somatosensory cortex, seen from three perspectives. **A.** External view of the left hemisphere, highlighting the location of S1 and S2, and adjacent posterior parietal cortex. (Note that “lateral sulcus” is an alternate name for Sylvian fissure.) **B.** Location of S1 and S2 as viewed in coronal section. **C.** Parasagittal cross section through central sulcus and postcentral gyrus, illustrating the four Brodmann areas that make up S1. One consequence of this organization, and the circuitry schematized here, is that there are more than one somatotopic map in S1. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

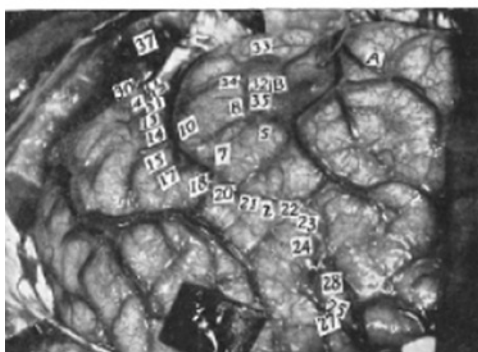
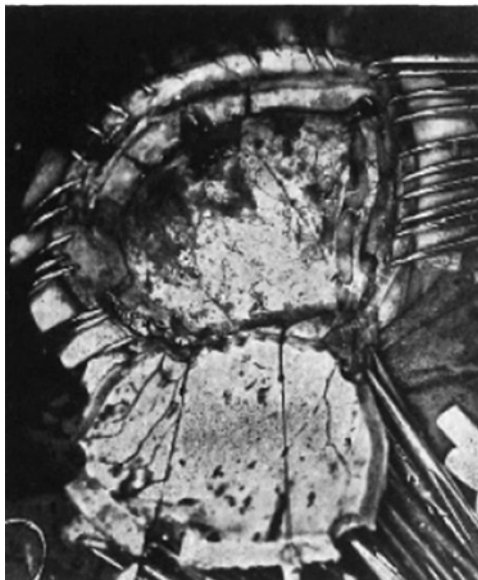


FIGURE 4.10 A surgical case from Penfield and Rasmussen (1950). The top panel illustrates the scalp of the patient, presurgery. The straight lines indicate where two of the cuts are to be made through the scalp, the “X” where the surgeon expects to find the “epileptogenic focus.” The middle panel shows an early stage of the surgery, at which the skull has been cut and bone flap folded away, but the dura is still intact. The bottom panel illustrates the exposed surface of the brain, with paper labels identifying locations where a response was

(Continued on page 104)

dictum, still by-and-large in force today, that *no brain is better than bad brain*. That is, if a region of cortex is so dysfunctional that its periodic episodes of uncontrolled activity lead to systemic (in this context, “brain-wide”) patterns of abnormal activity (i.e., seizures), the surgical removal of this **epileptogenic** tissue might bring an end to debilitating seizures, thereby allowing the remaining, healthy brain tissue to carry out its functions without interruption. An important corollary to this dictum, however, is that the surgeon does not want to damage cortex that supports functions that are critical for the patient’s quality of life. The procedure used to abide by this corollary has led to important discoveries about brain function.

Figure 4.10 illustrates the scalp and brain of an 18-year-old male whose **epilepsy** was determined, upon inspection of the exposed cortex, to stem from a tumor located in the right precentral gyrus. Thus, as Penfield and Rassmussen (1950, p. 6) wrote, “The surgical problem was to remove the tumor without damage to the precentral gyrus and thus to avoid paralysis” (Chapter 7 will detail the organization and function of primary motor cortex of the precentral gyrus). To accomplish this, they stimulated the surface of the brain at many locations (Figure 4.10). After performing many such surgeries, Penfield and Rassmussen (1950) were able to generate the “sensory **homunculus**” illustrated in Figure 4.11 (Tip 4.9). The precision of their measurements allowed them to document quantitatively the cortical magnification of various parts of the body, a feature analogous to the disproportionately larger cortical representa-

Tip 4.9

The word *homunculus* is Latin for “little man.” In psychology and neuroscience it is used in two contexts. Here, it refers to the representation of the human body as illustrated in Figure 4.11. It can also be used (often facetiously) in relation to the idea that a “little man inside our heads” is responsible for the conscious perception of sensory information, or for the coordination and control of brain activity.

tion of the central vs. peripheral visual field in V1. Indeed, recall from Chapter 3 that the fovea has a higher “packing density” of photoreceptors than do peripheral parts of the retina, and that foveal receptive fields are consequently much smaller than are peripheral receptive fields. Similarly, from Figure 4.11, we can deduce that the skin of the lower lip is much more densely innervated with sensory receptors than is the skin from the elbow, and that, consequently, cortical receptive fields representing the lower lip (and other parts of the face, and many parts of the hand ...) are much smaller than are cortical receptive fields representing the elbow (and the small of the back, and the knee ...). This disproportionate cortical representation of the surface of the body is perhaps most dramatically captured by a statue sculpted to preserve the cortical magnification determined by Penfield and Rassmussen (1950) which is viewable at Web Link 4.2.

FIGURE 4.10 (Continued) elicited by electrical stimulation. The responses at some of these locations are listed here. The voltage required to elicit each response is given in parentheses; words in quotes are verbal reports of the patient. 4 (1V) – Tingling, upper part of left leg; 7 (1V) – “Contraction of my left hand and arm.” There was closure of the hand and pronation of the forearm; 8 (1V) – Slight flexion of the hand. When asked, he said he had no idea what made him move his hand; 10 (1V) – Pronation of left forearm; 11 (1V) – Dorsiflexion of the left foot; 12 (1V) – Same; 13 (1V) – Sensation in back, down to the hips on the left; 14 (1V) – Sensation in face. When asked to indicate where, he pointed to the cheek and temple; 15 (1V) – Sensation in volar surface of forearm, slight movement of thumb; 22 (1V) – Sensation in index finger and thumb; 24 (1V) – Sensation in thumb; 25 (1V) – Sensation in tip of the tongue; 27 (1V) – Sensation in left side of tongue; 28 (1V) – Sensation in left lower lip; 30 (1V) – Sensation in big toe, left side. This cortex is adjacent to the falx; 32 (1V) – Patient reported contraction of his left leg like that before an attack. No movement, however, was observed; 33 (1V) – “A little of that feeling in the leg as though it were going to contract”; 34 (1½V) – Movement of left forearm and hand. When asked whether it was like an attack, he said, “Yes, but there is none of the beating sensation”; 37 (1½V) – Stimulation of the cortex 1cm below the surface in the longitudinal fissure produced contraction of the left leg with clonic movements. Source: Penfield and Rassmussen, 1950. Reproduced with permission of Macmillan.

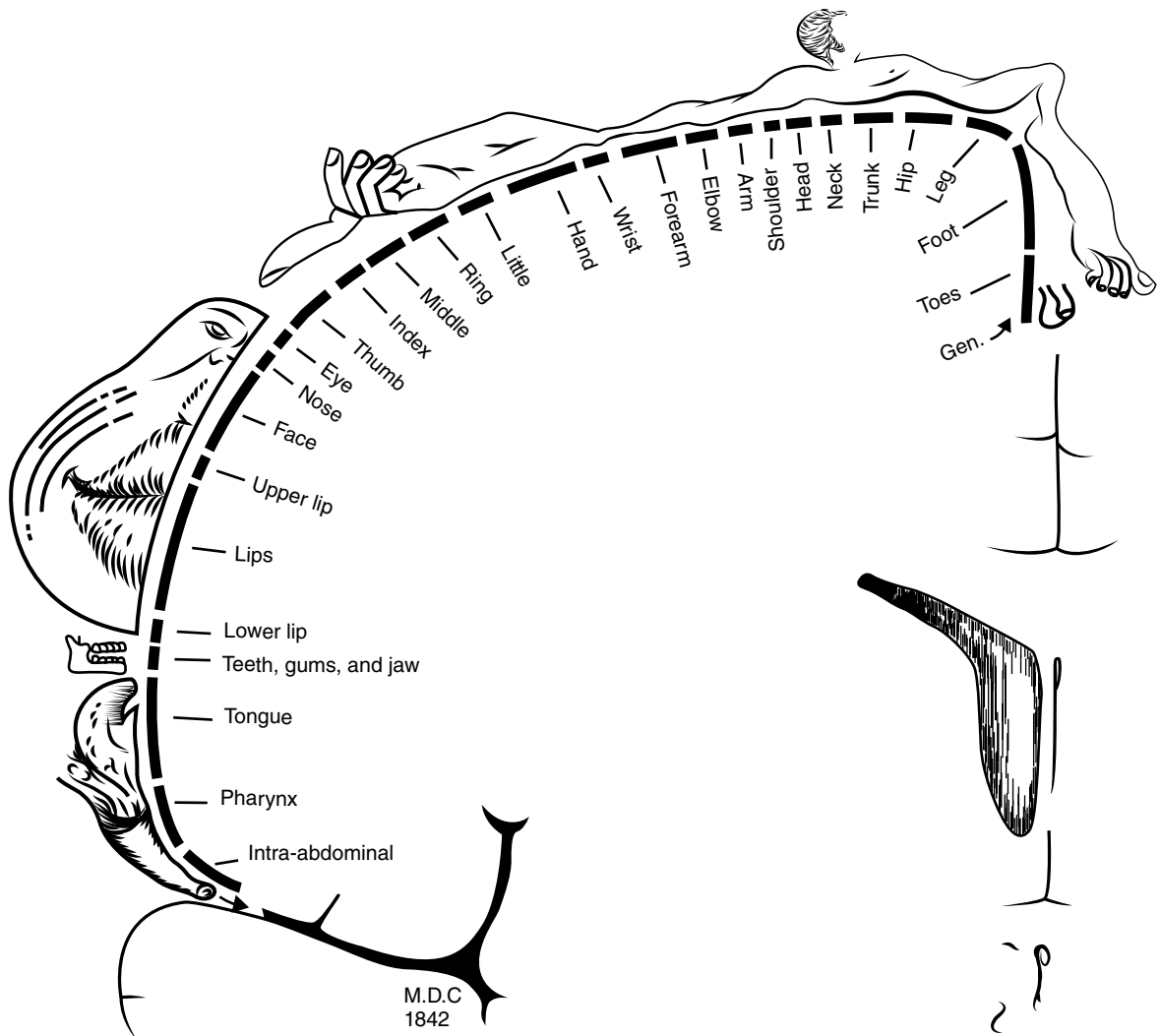


FIGURE 4.11 Depiction of the somatotopy of S1, as distributed along a paracoronal plane of section. The relative size of each body part reflects the surface area of cortex that represents it, as does the length of the line to which each label points. Three additional points are worthy of note. First, this scheme is a statistical composite derived from many individual cases, a fact that is evident if one tries to match up the location of labels and reports from the patient featured in *Figure 4.10* with the homunculus portrayed here. Second, this somatotopy can undergo rapid reorganization as a function of behavioral context or deafferentation (as will be described further along in this chapter, and is also demonstrated in the *Further Reading* by Kaas (2005)). Third, the discontinuities in the body representation of the somatosensory homunculus (e.g., the abrupt transition from face to fingers) underlie some fascinating neurological phenomena, also to be detailed further along in this chapter. Source: Penfield and Rassmussen, 1950. Reproduced with permission of Macmillan.

Somatotopy in the rodent: barrel cortex

This chapter will consider data from several species, and now we turn to the mouse, from which scientists have learned a great deal about somatosensation and,

more broadly, many principles of mammalian neural function. Many species of rodent, including the mouse and the rat, are nocturnal and/or spend much of their time underground. For such animals, somatosensation

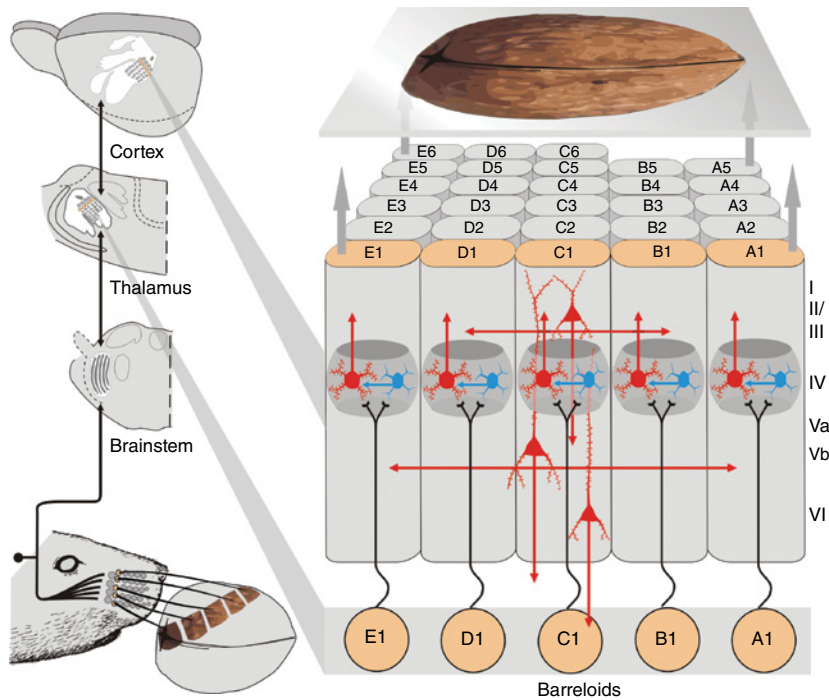


FIGURE 4.12 A diagram of the barrel system in a rodent. The column on the left illustrates a simplified rendition of the primary pathway from periphery to cortex. In the lower left, a cartoon rat is exploring an object, with the brown squares corresponding to the region of the object touched by each whisker. Each vibrissa follicle is innervated by mechanoreceptors that send their signals via the trigeminal nerve (of the PNS), in parallel, to one of many trigeminal nuclei in the brainstem (the first synapse of the system; only one nucleus illustrated here). Afferent fibers from the principal brainstem nucleus then project into the multiple “barreloids” of the VPM nucleus of the thalamus (second synapse), the arrangement of which mirror the topography of the animal’s face – that is, each barreloid receives information relating to just one whisker. Finally, thalamic afferents project into cortical area S1 (third synapse), which contains a map of the entire body surface of the animal (top left). Note that, as we have already seen with the human somatosensory homunculus, this map is distorted in a manner that corresponds to the innervation density of various parts of the animal’s body. The right side of the figure illustrates more detail of the thalamocortical projection and the cortex. Thalamic afferents project into layer IV, where both excitatory (red) and inhibitory (blue) neurons make up each barrel. Also illustrated for cortical column C1 (which corresponds to whisker C1 on the animal’s face) are pyramidal cells in superficial and deep layers of cortex. Floating fancifully “above” the barrel cortex is the artist’s rendition of the animal’s neural representation of the object being whisked (a walnut?), which is emergent from cortical processing in S1 and downstream areas. Source: Image provided by Dirk Schubert, Rolf Kötter, and Jochen Staiger.

via the whiskers is as important for navigation and exploration as is vision for humans. Indeed, an alert mouse actively explores its environment by “whisking,” a rhythmic back-and-forth sweeping of the whiskers that can vary in rate from 3 to 25 or more whisks per second. There is also considerable evidence that animals coordinate whisking rate and phase (i.e., where the whiskers are in the back-and-forth cycle) with movements of the head and paws. Whisking for a rodent can

be thought of as analogous to the near-constant eye movements that are characteristic of primates, something that we will consider in detail in *Chapter 8*.

The importance of whiskers to the rodent is reflected in the organization of its somatosensory cortex, in which one finds disproportionate representation of the snout, and within the “snout area,” the distinctive **barrel cortex** that processes information from the whiskers (*Figure 4.12*). The elegant simplicity

of the barrel system, including its “one-barrel-to-one-whisker” organization, makes it an ideal system for studying many principles of sensory processing. For example, there is considerable evidence that the phase of LFP oscillations within the barrel cortex is synchronized with the phase of whisking. The importance of such brain–body coordination is illustrated by

the study featured in *Research Spotlight 4.1*, in which optogenetic manipulation (*Methodology Box 4.2*) was used to causally demonstrate how the cortical processing of afferent sensory signals is much more effective when they are synchronized with gamma-band oscillations in S1. Importantly, the methods employed by this study provide insight into the mechanism for

METHODOLOGY BOX

4.2 Optogenetic manipulation of cellular activity

The logic of optogenetics in neuroscience can be described in four steps. (1) Clone a gene that codes for a protein that one wants to have expressed in a cell. (2) Insert this strand of DNA into a target population of cells (a process called transfection). (3) Wait for a suitable period of time for the protein to be synthesized and trafficked to the part of the cell where it will have its effects. Up until this point, the words written here could have also been used to describe gene therapy. The “opto-” part, however, is what makes this such a powerful experimental tool. In the case of the experiment of Cardin, Carlén et al. (2009; described in *Research Spotlight 4.1*), for example, the transfected DNA codes for channelrhodopsin-2 (ChR2), a cation channel activated by light of a very narrow wavelength centered at 470 nm. (The protein occurs naturally in the single celled algae *Chlamydomonas reinhardtii*, in which it serves the function of phototaxis – light-triggered movement.) Step 4, then, is to shine light of this wavelength on the cells containing the ChR2, thereby experimentally controlling when they spike. (Visit *Web Link 4.3* for a TED talk on optogenetics.)

Because Cardin, Carlén et al. (2009) were interested in controlling a specific type of neuron – so-called fast-spiking interneurons (“FS” in their figures) – they accomplished the transfection of the ChR2 gene with a viral vector that they knew would only penetrate into the nuclei of this particular kind of cell. *Figure 4.13* illustrates the optical control of these neurons. The experimental procedure was to first transfect FS neurons in rat barrel cortex with the ChR2 gene, then perform intracellular recordings of “regular spiking” (RS) pyramidal cells, as well as extracellular recordings of FS and RS neurons, while delivering brief pulses of 470 nm light onto this tissue via an optical fiber. **A.** Intracellular recording from an RS neuron shows a near-instantaneous inhibitory postsynaptic potential (IPSP) in response to a pulse of light. Note that because RS neurons were not transfected, the IPSP cannot have been caused directly by the light. Rather, it must have been the result of the release of GABA from transfected RS neurons onto this pyramidal cell (i.e., the pulse of light produced an action potential in one or more RS neurons). **B.** The latency between pulses of light delivered at low or high intensity, and two measures: action potentials in FS neurons (filled symbols) and IPSPs in RS neurons (unfilled symbols). These data indicate that the propagation of the action potential in an interneuron, the resultant release of GABA, and the consequent postsynaptic influx of Cl^- ions all transpire in less than 1 ms. **C.** A functional consequence of sustained optical stimulation of FS inhibitory interneurons is to block sensory-evoked responses in RS neurons to a whisker deflection. Left-hand panel shows sustained firing of an FS cell in layer 2/3 in response to a 10 ms light pulse (blue line). Center panel shows the response of an adjacent RS cell to a whisker deflection (red bar). Right-hand panel shows that when light is delivered coincident with whisker deflection the result is to suppress activity in the RS neuron to below its baseline level.

this coordination, by establishing a causal role for inhibitory interneurons within a barrel for controlling the dynamics of gamma-band oscillations in the LFP. The potential implications for this kind of work are manifold. Because the barrel is part of a specialized type of cortical column, what's learned about its functioning is likely to generalize to cortex in many other

Tip 4.10

The author's intent here is not to be pejorative. After all, many of his studies have been based on “merely observing” correlations between an independent and a dependent variable. Rather, it is to highlight the inferential advantage to be gained when the system being studied can be exogenously perturbed.

Tip 4.11

To definitively establish that a particular neuron is an inhibitory interneuron, an excitatory neuron that projects only locally within a cortical column, a “long-distance” pyramidal neuron, or some other type of cell, requires a combination of histological and biochemical information that is not always practical, or possible, to acquire in a physiological experiment. Thus, in papers making a point about different physiological responses between putatively different types of neuron, it's not uncommon to find authors limiting their descriptions to the properties that they can directly observe (such as “fast-spiking” (FS) vs. “regular-spiking” (RS)), even though “everybody knows” that what they are describing is interneurons vs. excitatory neurons.

RESEARCH SPOTLIGHT

4.1 Cardin, Carlén et al. (2009), optogenetic exploration of gamma-band oscillations in somatosensory cortex

In *Chapter 2*, the phenomenon of oscillatory synchrony was introduced as a possible mechanism for organizing neural activity, such as the selection of to-be-emphasized information. The study highlighted here is noteworthy because it provides *causal*, mechanistic insight into how one type of neuronal oscillation – gamma-band oscillations – is generated. Further, it demonstrates the importance of these oscillations for the cortical processing of afferent sensory signals. It accomplishes this by using optogenetic methods to exogenously control gamma-band oscillations in one class of neurons located in barrel cortex – fast-spiking (FS) inhibitory interneurons. In so doing, it illustrates how experimental methods that exert external control over an aspect of the brain or brain function can support stronger inferences than would have an experiment that “merely observed” the kind of activity that is described here (*Tip 4.10*).

Cardin, Carlén et al. (2009) used the methods outlined in *Methodology Box 4.2* to explore the hypothesis that inhibitory interneurons of barrel cortex (the “blue” neurons from *Figure 4.12*; the “FS” neurons from *Methodology Box 4.2*) govern the gamma-band oscillations previously observed in LFPs recorded from barrel cortex (*Tip 4.11*). The experiment proceeded in three stages. First, as illustrated in *Figure 4.13*, the authors established that their preparation gave them (a) direct optical control over these FS neurons and, by extension, (b) indirect inhibitory control over excitatory cells (the “red” neurons from *Figure 4.12*; the “RS” neurons from *Methodology Box 4.2*). Second, they established that FS cells have a preferred oscillatory frequency in the 40–50 Hz range (*Figure 4.14*). Third, they implemented a test of

Research Spotlight 4.1 Continued

the idea that “gamma oscillations ... have a functional impact on cortical information processing by synchronizing the output of excitatory neurons” (p. 666). To do so, they imposed an oscillation at this frequency by stimulating FS cells with light pulses delivered at 40 Hz, and recorded the responses of RS cells to deflections of a whisker performed at five different phases relative to the gamma cycle (Figure 4.15.A). The logic was that, if gamma-band oscillations serve a **gating** function, one should see optimal responses to afferent sensory information by layer-4 excitatory neurons when that information

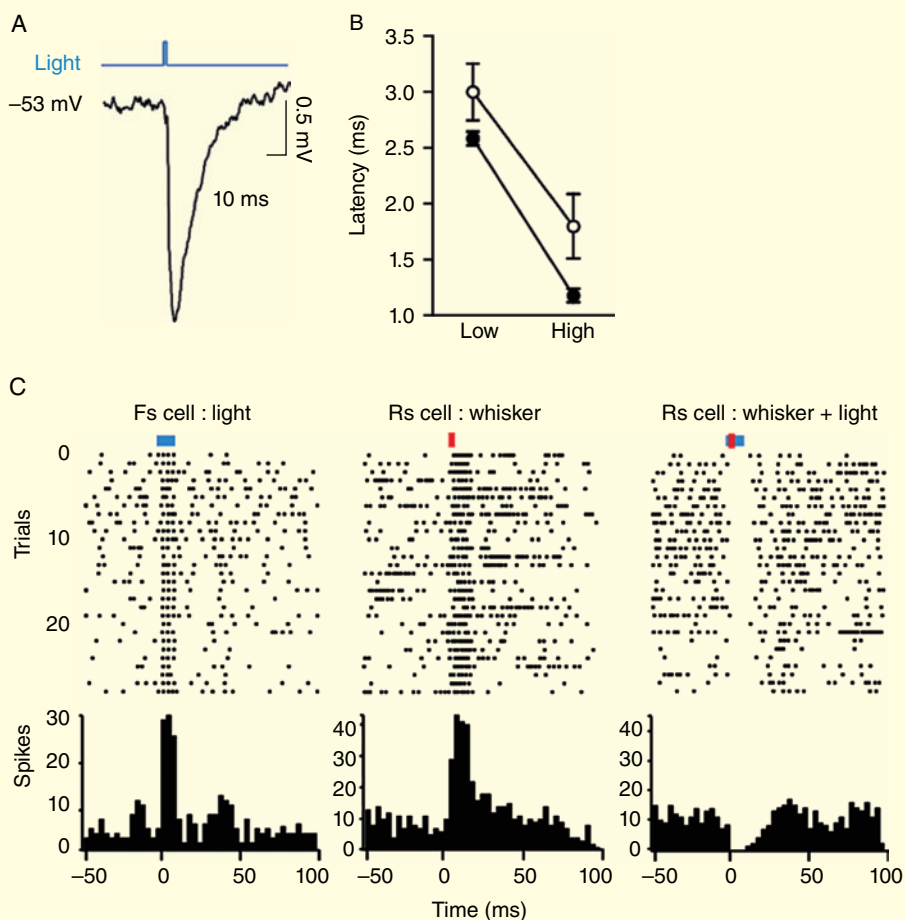


FIGURE 4.13 The effects of optical control of fast-spiking (FS) interneurons. See *Methodology Box 4.2* for an explanation of this figure. Source: Adapted from Cardin, Carlén, Meletis, Knoblich, Zhang, Deisseroth, Tsai, and Moore, 2009. Reproduced with permission of Nature Publishing Group.

Research Spotlight 4.1 Continued

arrives at the optimal time relative to the ongoing gamma-band oscillation in this neuron's membrane potential (*Figure 4.15.B*). The results, illustrated in *Figure 4.15.C–E*, demonstrated the predicted gating function of gamma-band oscillations, in that the response to whisker stimulation of a layer-4 RS neuron was sensitive to the phase of the optically entrained LFP. That is, unless the whisker stimulation was delivered to coincide with the peak of the LFP, it lacked the magnitude of response (in spikes per trial) of the baseline response.

Why weren't optimally timed whisker stimulations stronger than the baseline? One possible answer comes from looking back at *Figure 4.12*, and noting that thalamocortical axons carrying sensory signals from barreloids to barrels send collaterals that synapse directly on layer-4 interneurons. This suggests that the circuitry is in place to coordinate interneuron activity with the processing of the afferent sensory signal, such that on “baseline” trials (when no optical stimulation is being delivered), the circuitry naturally produced an optimal response.

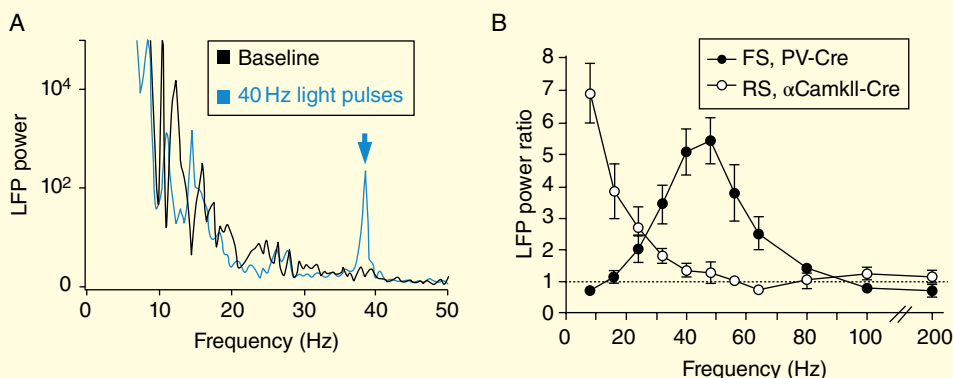


FIGURE 4.14 FS cells have a preferred oscillatory frequency in the 40–50 Hz range. **A** shows a power spectrum of the LFP in barrel cortex as recorded in the absence of optical stimulation (“Baseline”; black) and during the delivery of 40 Hz light pulses (blue). **B** shows that, across a wide range of frequencies of optical stimulation, the LFP power ratio ($\text{LFP}_{\text{optical stimulation}} / \text{LFP}_{\text{baseline}}$) for transfected FS neurons (filled symbols) is highest in the 40–50 Hz range, suggesting that this is the “naturally preferred” frequency (or **resonant frequency**) of these cells. For comparison, in a different group of animals, RS cells (unfilled symbols) were transfected with the ChR2 gene. The highest LFP power ratio in these animals, and thus the presumed resonant frequency of RS cells, was at 8 Hz (i.e., in theta frequency band). Source: Adapted from Cardin, Carlén, Meletis, Knoblich, Zhang, Deisseroth, Tsai, and Moore, 2009. Reproduced with permission of Nature Publishing Group.

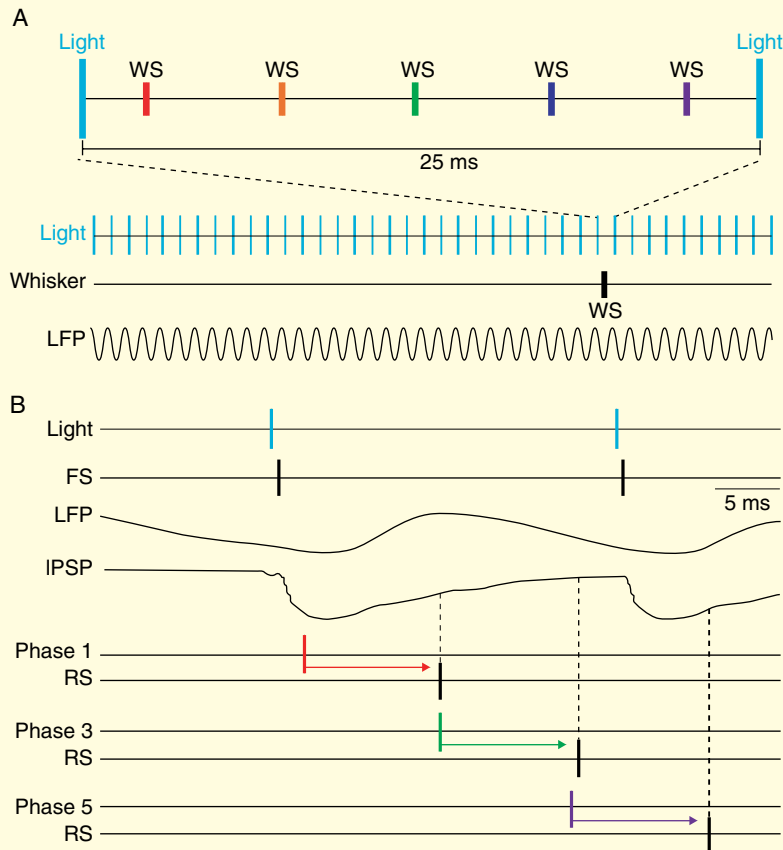
Research Spotlight 4.1 *Continued*

FIGURE 4.15 Optogenetically controlled gamma-band oscillations influence cortical processing of afferent sensory signals. **A.** Gamma-band oscillation is imposed by delivering light pulses at 40 Hz to ChR2-transfected FS cells (*Methodology Box 4.2*), and whisker deflections (WS) are performed at five different phases relative to the gamma cycle. **B.** Schematic model of the expected effects of this procedure: The top trace shows the once-every-25 ms (i.e., 40 Hz) pulses of light; the next, the consequent action potential in FS neurons; the next, the resultant gamma-band oscillation of the LFP; and the next, the concurrent fluctuations in membrane potential experienced by a pyramidal cell ("IPSP"). An afferent signal from a WS delivered at phase 1 (red WS in panel A) would arrive at the RS cell at the point of maximal inhibition in the gamma-phase, and would produce a spike roughly 10 ms later; a WS delivered at phase 3 (green) would arrive at a point of rising membrane potential, and produce a spike near the peak membrane potential; and a WS delivered at phase 5 (purple) would arrive near the peak of the membrane potential, the resultant spike occurring near the trough of the membrane potential. **C.** Responses of a layer 4 RS neuron to WSs in the absence of optical stimulation (baseline). **D.** Responses of the same neuron to WSs delivered at each of the five phases portrayed in A. **E.** Average spikes/trial for baseline (B) and each of the five phases. Source: Adapted from Cardin, Carlén, Meletis, Knoblich, Zhang, Deisseroth, Tsai, and Moore, 2009. Reproduced with permission of Nature Publishing Group.

Research Spotlight 4.1 *Continued*

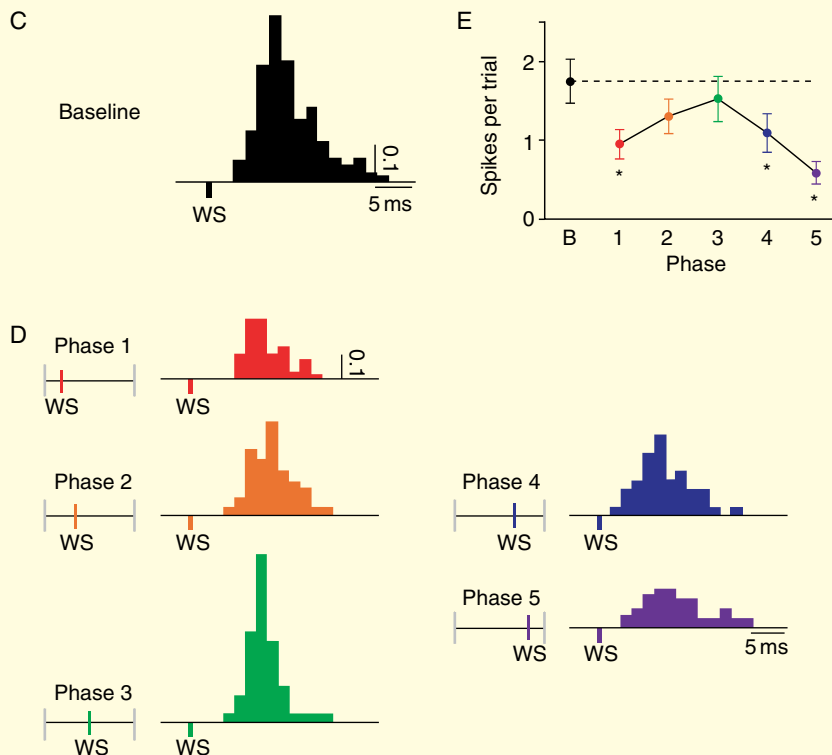


FIGURE 4.15 (Continued)

brain regions, and in many other species. Further, because there is compelling correlational evidence that human visual perception is influenced by the coordination of timing between stimulus onset and oscillatory phase of the occipital EEG (see *Further Reading* by Van Rullen and Dubois, 2011), the findings detailed in *Research Spotlight 4.1* may offer important insights into the functioning of human sensory systems, including vision.

Somatosensory plasticity

Plasticity refers to the ability of the nervous system to change its function, and sometimes its structure, in response to a change in the internal or external envi-

ronment. The somatosensory system, including rodent barrel cortex, is a system in which a great deal of research has been conducted on the principles and mechanisms underlying plastic change in the brain. For example, many studies have been performed in which an individual whisker is clipped close to the skin, so that it no longer comes in contact with objects during whisking. This results in a reorganization of the representation of this, and adjacent, whiskers in barrel cortex. Such research has provided important insights into developmental, genetic, molecular, and pharmacological factors underlying experience-dependent plasticity (*Web Link 4.4* collects many references to this research). Here, however, we turn to systems-level research in the nonhuman

primate to examine the phenomenon of somatosensory plasticity, and some of the principles underlying it. As we shall see, this research sheds light on fascinating (human) neurological syndromes that can result from damage to the somatosensory periphery.

Use-dependent plasticity

Some of the pioneering work in this area has been carried out by Michael Merzenich and colleagues at the University of California, San Francisco. One of the motivations for these studies was an observation made while mapping the cortical representation of fingers within the hand area: the surface area of cortex representing a given finger could differ dramatically from one animal to the next. This gave rise to the idea that cortical representation of the body might not be rigidly fixed by a genetically determined anatomical structure, but may, instead, be a dynamic property that, even in the adult, can change to reflect changes in experience (i.e., two monkeys might have differently organized hand areas because they use their hands differently.) To test this idea, Merzenich and colleagues devised experiments in which the somatosensory experience of the animal was altered in systematic ways, and the cortical representation of the implicated body part was examined for evidence of experience-dependent change. One type of study involved training the animal to make discriminations between vibrations of different frequencies delivered to one specific location on one finger (*Figure 4.16*). Results indicated the cortical representation of the targeted area grew dramatically after extensive training. Importantly, this effect was specific to the piece of skin that directly experienced this stimulation, because “before and after” mapping of the cortical representation of the contralateral finger tip, which hadn’t been trained, showed no change. Additionally, this plasticity was shown to be sensitive to cognitive factors, because little change would occur in the cortical representation of a finger that received the same amount of vibrational stimulation as the one illustrated in *Figure 4.16*, but did so while the monkey was paying attention to a difficult auditory discrimination task. The fact that attention to tactile stimuli seems to be necessary for somatosensory plasticity to occur is yet another demonstration of a critical role for top-down input onto primary sensory cortex.

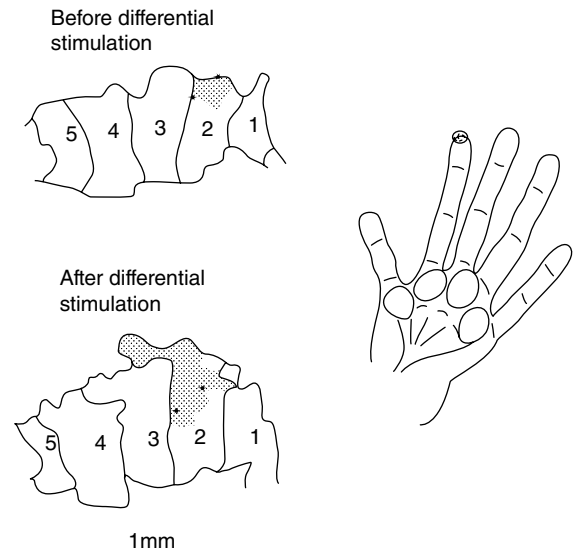


FIGURE 4.16 Somatosensory reorganization after training. The upper-left figure shows the representation of the five fingers of the hand (thumb = “1”; index finger = “2”; etc.) in primary somatosensory cortex of area 3B (see *Figure 4.17*) of an adult owl monkey, with the shaded portion indicating the location of neurons with receptive fields corresponding to the tip of the index finger (drawing on right). The bottom-left figure shows the dramatic increase in the number of neurons with this same receptive field, the asterisks indicating the same anatomical location in both images. Note that this reorganization occurs primarily within the preexisting representation of digit 2, with the expansion at the top of the image corresponding to a rostral “invasion” of area 3A. The boundaries between finger zones, however, are largely unchanged. Source: Adapted from Merzenich, Recanzone, Jenkins, Allard, and Nudo, 1988. Reproduced with permission of Wiley.

The brain mimics the body

A second type of study, illustrated in *Figure 4.17*, involved surgically fusing the skin of two fingers in order to transform them into effectively a single, “fat” finger in the middle of the hand. Here, the result was that the discrete, discontinuous boundary between the two fingers, which had existed prior to the surgery (and is illustrated in *Figure 4.17*), was replaced by a “blurred” cortical representation that did not discriminate between the two previously separate fingers. This is a concrete example of a principle that we will

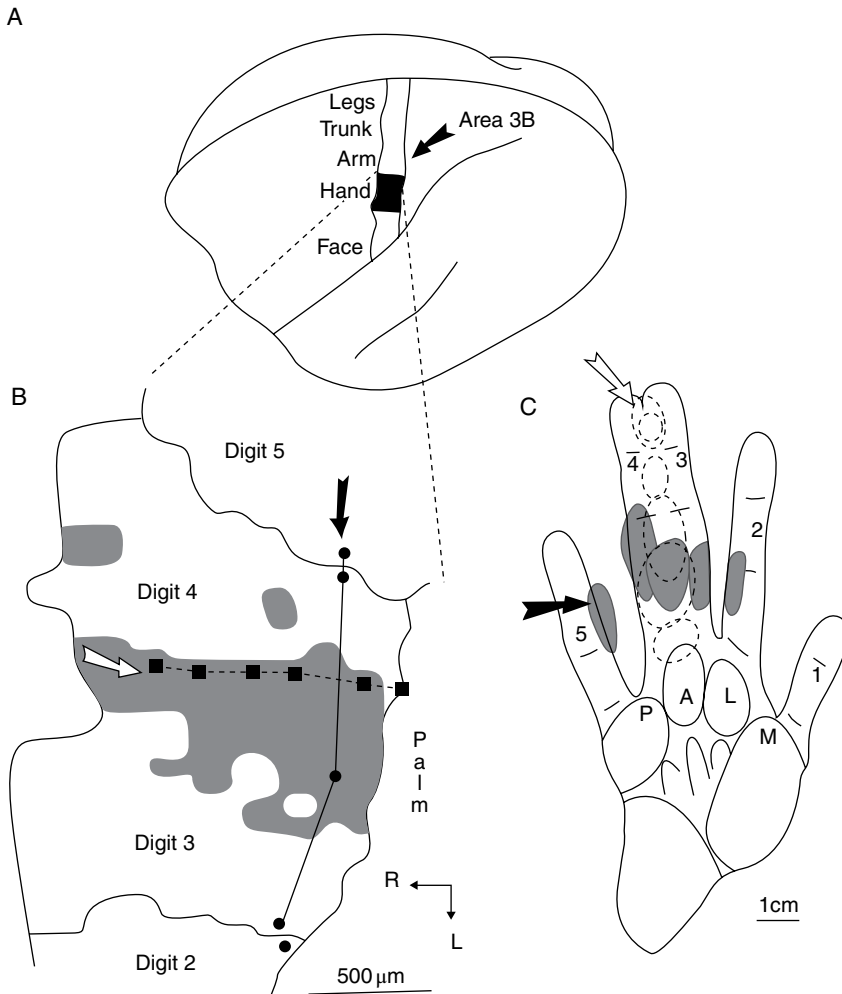


FIGURE 4.17 Somatosensory reorganization after a change to the body. Prior to surgical fusion of digits 3 and 4 (panel **C**), no neurons were found that responded to tactile stimulation of both fingers. 5½ months after the peripheral change, however, many such neurons were found, covering the cortical regions that are shaded in panel **B**. Additionally, the abrupt boundary that had existed between the cortical representation of these two digits could no longer be identified. The white arrow in panel B shows the path and direction of a rostral-to-caudal electrode penetration, with each square along the path corresponding to each of the dotted-line receptive fields drawn onto panel C (i.e., starting at tips of 3–4 and descending to the palm). The black arrow in panel B shows the path and direction of a medial-to-lateral penetration, with each circle along the path corresponding to the shaded receptive fields drawn onto panel C. Source: Clark, Allard, Jenkins, and Merzenich, 1988. Reproduced with permission of Nature Publishing Group.

consider in some depth in *Chapter 10*: neural plasticity can be construed as the process of the brain changing its structure so as to more closely model the structure of the world that it is representing. In the case

illustrated in *Figure 4.17*, the monkey's physical reality was that a hand that previously had five independent fingers now had four. The consequent plastic reorganization in somatosensory cortex can be seen as its

brain reshaping itself to better represent this new state of affairs. In *Chapter 7*, we shall see that analogous plasticity in the motor system is characteristic of learning new motor skills.

Mechanisms of somatosensory plasticity

In the demonstrations of cortical plasticity that we have considered in the previous two paragraphs, there are at least two possible explanations for how the cortical representation of a part of the body increases in size (*Figure 4.16*) or merges with another (*Figure 4.17*). One is that afferent axons that drove one discrete region of cortex have “sprouted” into the new cortical territory (e.g., by sending newly sprouted axon collaterals into the “conquered” region). A second is that previously present, but for some reason quiescent, axon collaterals representing the expanding part of the body were “unmasked” as a result of the changes inducing the plasticity. Evidence for the latter comes from amputation/transection studies in which either a part of the body (e.g., a finger), or a sensory nerve innervating that part of the body, is experimentally severed. In these situations, the expansion of representation of adjacent, undamaged skin surfaces (e.g., the adjacent finger) is evident almost immediately, a timescale much too fast for sprouting (which would require new protein synthesis, the growth of the axon collateral, etc.). Thus, unmasking of preexisting axonal terminals is the most plausible explanation for at least the earliest stages of cortical plasticity. By what mechanism was the now-influential axon previously “masked”? One explanation is by lateral inhibition driven by the previously dominant excitatory input, but alternative models exist.

One take-away from this section is that, although illustrated depictions of somatotopic representation in S1 suggest a static homunculus, the cortical representation of the body’s surface may, instead, be quite dynamic. The anatomical substrate that affords rapid reorganization in somatosensory cortex may be somewhat different from the one-to-one projection zones suggested by the “one-whisker-one-barrel” scheme for rodent barrel cortex. Instead, a manifold of overlapping afferent inputs, the majority of which are typically “masked” due to local inhibitory dynamics, may be the norm, at least in the primate brain. (Having read through this section, you’ll now have more context for *Web Clip 4.1*, on “plasticity in the auditory cortex.”)

Phantom limbs and phantom pain

Although the type of experiment described in the preceding paragraph may seem particularly sadistic from the standpoint of the experimental animal, such studies have given important insight into human neurological disorders, such as the “phantom limb” syndrome that can be experienced by patients who have either lost a limb or suffered damage to the spinal cord such that afferent sensory information from that limb no longer reaches the brain. These patients often report experiencing tactile sensations, and sometimes pain, that feel as though they are coming from the limb that they have lost. Although it was long assumed that these sensations were due to scrambled firing in nerve fibers at the stump of the lost limb, this explanation has been ruled out by the fact that injecting local anesthetic into the stump, which silences these fibers, does not quiet the phantom sensations. A synthesis of much of the research presented in this chapter, however, suggests an alternative: it may be that plastic reorganization of the cortex, resulting from the loss of innervation from the amputated limb, results in cortical signals that get “misinterpreted” as coming from the no-longer-existing limb.

A hypothesized mechanism

Evidence for a cortical-reorganization account of phantom sensations comes from patients such as the man in *Figure 4.18*, whose left arm had been amputated four weeks previously. In a deceptively simple but enormously influential case study, the neurologist and cognitive neuroscientist V. S. Ramachandran brushed several parts of this patient’s body with a cotton swab. Only when two parts of the body were brushed – the side of the face ipsilateral to the amputation, and the upper arm above the amputation site – did the patient report two simultaneous tactile sensations: one on the skin that was actually touched, and one on a part of the amputated hand. Ramachandran argued that the location of these “reference zones” was not arbitrary. Recall from the “Penfield homunculus” (*Figure 4.11*) that the cortical representation of the hand is located immediately dorsal to that of the face, and immediately ventral to that of the arm. Thus, the region of cortex that previously represented this patient’s left hand may have been “invaded” by afferents carrying information



FIGURE 4.18 The reference zones of the “phantom” index finger (I), pinky (P), thumb (T), and ball of the thumb (B) of the left hand, on the face of a left-arm amputee. Source: Ramachandran and Hirstein, 1998. Reproduced with permission of Oxford University Press.

about the face, as well as about the upper arm. More specifically, from *Figure 4.18*, one can infer that afferent signals from the left cheek may now be driving cortical tissue that had previously represented the left thumb, that afferent signals from the left upper lip are now driving cortical tissue that had previously represented the left index finger, and so on. The phantom sensations experienced by the patient, then, may indicate that the reference zones illustrated in *Figure 4.18* may, in fact, correspond to newly formed receptive fields of neurons occupying what had been the hand area.

Helpful or harmful?

Ramachandran’s experiments, and follow-up work by his group and others, has sparked enthusiastic discussion and debate, among philosophers as well as cognitive neuroscientists. One unresolved problem posed by the cortical-reorganization model is that of reconciling the beneficial, indeed necessary, functions of plasticity with the obviously undesirable, and sometimes debilitating, clinical phenomena of

phantom limbs and phantom pain. How is it that a capacity that serves us so well when it comes to learning to read Braille or learning a musical instrument can also produce distracting, and sometimes excruciatingly painful, sensations that seem to be coming from a limb that no longer exists? A second relates to the conscious perception of touch. The cortical-reorganization model of the phantom-limb phenomenon must assume that the read-out of signals from the cortex that had represented the hand is still being “interpreted” as corresponding to the hand. Where and how does this interpretation take place? And why can’t this erroneous interpretation be overridden by the patient’s simultaneous awareness that the hand from which s/he seems to be experiencing sensations no longer exists?

Proprioception

Before concluding this section, we need to introduce the somatosensory modality of **proprioception**, which refers to the sense of where different

parts of the body are relative to one another. Think back for a moment to the thought problem posed in *Tip 4.6*. One need not understand the nuts and bolts of how sound localization works to understand the following: a sound originating 10 feet (3 meters) behind you will produce markedly different IIDs and ITDs depending on whether your head is pointed straight ahead or whether it is turned to the side (e.g., if you are looking off to your left); yet the proverbial tiger that is stalking you is in the same location relative to your body, and so your most appropriate response is the same regardless of how your head happens to be oriented at the moment. The same is true for vision: if you are looking straight ahead, the tiger that is directly in front of you would occupy your foveal vision; whereas if your head is turned to the left, the tiger would occupy the extreme right side of your visual field. Thus, to accurately process information coming in through the ears, or through the eyes, the brain needs to know how the head is positioned relative to the torso. (And for vision, an additional factor that will receive more attention in *Chapter 8* is that the brain also needs to know the position of the eyes relative to the head.) Information about the position of the head relative to the torso is

provided by proprioceptive receptors located in the neck. For our purposes, these are stretch receptors that gauge the state of the neck muscles.

Adieu to sensation

The final section of *Chapter 3* already disabused us of the idea that sensation and perception can be construed as discrete, serial processes, the latter “taking over” *where* and *when* the former terminates. This was reinforced in the present chapter, via evidence for top-down control of the very process of transduction of the acoustic signal in the auditory system, and for the synchronization of whisking with the oscillatory state of somatosensory cortex. Beginning with *Chapter 5*, however, we will be fully “in the brain,” and squarely in the domain of perception. Or will we? Those who prefer their science to come in tidy, clearly demarcated superordinate categories may be dismayed to learn that we’re transitioning out of the frying pan of the interactive continuum along which sensation and perception lie, and into the fire of an interactive continuum linking perception and memory. It’ll be messy at times. We’re going to get dirty. But it will be so worth it!

END-OF-CHAPTER QUESTIONS

1. How might the mechanical properties of the cochlea result in the fact that the tuning of auditory nerve fibers broadens with increasing sound intensity?
2. What is meant by the term “spectral decomposition”? How is the operation carried out in the auditory system?
3. What components of the AEP are sensitive to attentional state? What can one infer from this about which stages of auditory processing are, vs. are not, influenced by attention?
4. Which part of your body is more densely innervated with mechanoreceptors, your tongue or your wrist? What are the corresponding differences in how these two parts of your body are represented in S1?
5. What are two properties of optogenetics that make it such a powerful tool for studying the brain?
6. How specific or nonspecific is use-dependent plasticity? How dependent on attention? Why are the answers to these questions important?
7. What kind of data have been used to assess the contributions of unmasking vs. sprouting to somatosensory reorganization after deafferentation? What might be a broader implication to this question?
8. Does the cortical-reorganization hypothesis of phantom-limb syndrome assume that a neuron can simultaneously have two receptive fields? Explain.

REFERENCES

- Belluck, Pam. 2013. "Could Hearing Loss and Dementia Be Connected?" *New York Times*, February 11. http://www.nytimes.com/2013/02/12/science/could-hearing-loss-and-dementia-be-connected.html?ref=disabilities&_r=0. Interview with Katherine Bouton.
- Bouton, Katherine. 2013. "Straining to Hear and Fend Off Dementia." *New York Times*, February 11. <http://well.blogs.nytimes.com/2013/02/11/straining-to-hear-and-fend-off-dementia/>.
- Brugge, John F., and Matthew A. Howard III. (2002). "Hearing." In *Encyclopedia of the Human Brain: Vol. 2* (pp. 429–448). San Diego: Academic Press.
- Cardin, Jessica A., Marie Carlén, Konstantinos Meletis, Ulf Knoblich, Feng Zhang, Karl Deisseroth, ... Christopher I. Moore. 2009. "Driving Fast-Spiking Cells Induces Gamma Rhythm and Controls Sensory Responses." *Nature* 459 (7247): 663–667. doi: 10.1038/nature08002.
- Clark, Sharon A., Terry Allard, William M. Jenkins, and Michael M. Merzenich. 1988. "Receptive Fields in the Body-Surface Map in Adult Cortex Defined by Temporally Correlated Inputs." *Nature* 332 (6163): 444–445.
- Johnson, Jeffrey S., Bornali Kundu, Adenauer G. Casali, and Bradley R. Postle. 2012. "Task Dependent Changes in Cortical Excitability and Effective Connectivity: A Combined TMS-EEG Study." *Journal of Neurophysiology* 107 (9): 2383–2392. doi: 10.1152/jn.00707.2011.
- Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell, eds. 2000. *Principles of Neural Science*. New York: McGraw-Hill.
- Lin, Frank R., E. Jeffrey Metter, Richard J. O'Brien, Susan M. Resnick, Alan B. Zonderman, and Luigi Ferrucci. 2011. "Hearing Loss and Incident Dementia." *Archives of Neurology*, 68 (2): 214–220. doi: 10.1001/archneurol.2010.362.
- Lin, Frank R., Kristine Yaffe, Jin Xia, Qian-Li Xue, Tamara B. Harris, Elizabeth Purchase-Helzner, ... Eleanor M. Simonsick, for the Health ABC Study Group. 2013. "Hearing Loss and Cognitive Decline in Older Adults." *JAMA Internal Medicine* 173 (4): 293–299. doi: 10.1001/jamainternmed.2013.1868.
- Merzenich, Michael M., Gregg H. Recanzone, William M. Jenkins, Terry T. Allard, and Randolph J. Nudo. 1988. "Cortical Representational Plasticity." In *Neurobiology of Neocortex: Report of the Dahlem Workshop on Neurobiology of Neocortex Berlin, 1987 May 17–22*, edited by Pasko Rakic and W. Singer, 41–67. Chichester: John Wiley and Sons.
- Penfield, Wilder, and Theodore Rasmussen. 1950. *The Cerebral Cortex of Man. A Clinical Study of Localization of Function*. New York: Macmillan.
- Picton, Terrence W., Steven A. Hillyard, Howard I. Krausz, and Robert Galambos. 1974. "Human Auditory Evoked Potentials. I: Evaluation of Components." *Electroencephalography and Clinical Neurophysiology* 36: 179–190. doi: 10.1016/0013-4694(74)90155-2.
- Ramachandran, V. S., and William Hirstein. 1998. "The Perception of Phantom Limbs: The D. O. Hebb Lecture." *Brain* 121 (9): 1603–1630. doi: 10.1093/brain/121.9.1603.
- Sekuler, Robert, and Randolph Blake. 1985. *Perception*. New York: Alfred A. Knopf.
- Wolfe, Jeremy M., Keith R. Kluender, Dennis M. Levi, Linda M. Bartoshuk, Rachel S. Herz, Roberta Klatzky, ... Daniel M. Merfeld. 2012. *Sensation and Perception*, 3rd ed. Sunderland, MA: Sinauer Associates.
- Yantis, Steven. 2014. *Sensation and Perception*. New York: Worth.

OTHER SOURCES USED

- Ganguly, Karunesh, and David Kleinfeld. 2004. "Goal-Directed Whisking Increases Phase-Locking between Vibrissa Movement and Electrical Activity in Primary Sensory Cortex in Rat." *Proceedings of the National Academy of Sciences of the United States of America* 101 (33): 12348–12353. doi: 10.1073/pnas.0308470101.
- Morel, Anne, Preston E. Garraghty, and Jon H. Kaas. 1993. "Tonotopic Organization, Architectonic Fields, and Connections of Auditory Cortex in Macaque Monkeys." *Journal of Comparative Neurology* 335 (3): 437–459. doi: 10.1002/cne.903350313.
- Picton, Terrence, and Steven A. Hillyard. 1974. "Human Auditory Evoked Potentials. II: Effects of Attention." *Electroencephalography and Clinical Neurophysiology* 36: 191–199. doi: 10.1016/0013-4694(74)90156-4.
- Recanzone, Gregg H., Michael M. Merzenich, William M. Jenkins, Kamil A. Grajski, and Hubert R. Dinse. 1992. "Topographic Reorganization of the Hand Representation in Cortical Area 3b of Owl Monkeys Trained in a Frequency-Discrimination Task." *Journal of Neurophysiology* 67 (5): 1031–1056.
- Schubert, Dirk, Rolf Kötter, and Jochen F. Staiger. 2007. "Mapping Functional Connectivity in Barrel-Related Columns Reveals Layer- and Cell Type-Specific Microcircuits." *Brain Structure and Function* 212 (2): 107–119.

- Woolsey, Thomas A., and Hendrik Van der Loos. 1970. "The Structural Organization of Layer IV in the Somatosensory Region (SI) of Mouse Cerebral Cortex: The Description of a Cortical Field Composed of Discrete Cytoarchitectonic Units." *Brain Research* 17 (2): 205–242. doi: 10.1016/0006-8993(70)90079-X.

FURTHER READING

- Belluck, Pam. 2013. Interview with Katherine Bouton. "Could Hearing Loss and Dementia Be Connected?" *New York Times*, February 11. http://www.nytimes.com/2013/02/12/science/could-hearing-loss-and-dementia-be-connected.html?ref=disabilities&_r=0.
An interview with author Katherine Bouton, whose column in The New York Times is also referenced here.
- Bouton, Katherine. 2013. "Straining to Hear and Fend Off Dementia." *New York Times*, February 11. <http://well.blogs.nytimes.com/2013/02/11/straining-to-hear-and-fend-off-dementia/>.
Newspaper column from a 65-year-old woman that eloquently captures the difficulties in daily life, as well as broader concerns, of a person experiencing age-related hearing loss.
- Heckert, Justin. 2012. "The Hazards of Growing Up Painlessly." *New York Times Magazine*, Nov 18, 27–31, 54–55, 4.
An in-depth look at the life-altering consequences of being physiologically unable to experience a somatosensory modality that most of us take for granted, and often try to suppress.
- Kaas, Jon H., ed. 2005. *The Mutable Brain: Dynamic and Plastic Features of the Developing and Mature Brain*. Amsterdam: Taylor and Francis.
A definitive collection of chapters on the plasticity of cortex, including such topics as time course of reorganization, and context dependent expansion/contraction of sensory and motor representations.
- Luck, Steven J. 2005. *An Introduction to the Event-related Potential Technique*. Cambridge, MA: MIT Press.
An accessible, authoritative introduction to ERPs, the reading of which will make one an informed and critical consumer of ERP research, regardless of whether or not one uses the technique.
- Padden, Carol, and Tom Humphries. 2005. *Inside Deaf Culture*. Cambridge, MA: Harvard University Press.
A very compelling "insider's account" of what it's like to be deaf in contemporary society, written by two professors of communication at the University of California, San Diego.
- Ramachandran, V. S., and William Hirstein. 1998. "The Perception of Phantom Limbs: The D. O. Hebb Lecture." *Brain* 121 (9): 1603–1630. doi: 10.1093/brain/121.9.1603.
A review from this celebrated neurologist and cognitive neuroscientist on the history, phenomenology, and neurobiology of this fascinating disorder.
- Van Rullen, Rufin, and Julien Dubois. 2011. "The Psychophysics of Brain Rhythms." *Frontiers in Psychology* 2 (203): 1–10. doi: 10.3389/fpsyg.2011.00203.
A review of recent research on the idea that many perceptual and attentional processes may be regulated by neuronal oscillations.
- Wolfe, Jeremy M., Keith R. Kluender, Dennis M. Levi, Linda M. Bartoshuk, Rachel S. Herz, Roberta Klatzky, ... Daniel M. Merfeld. 2012. *Sensation and Perception*, 3rd ed. Sunderland: Sinauer Associates.
Chapters 14 and 15 of this textbook provide excellent introductions to the neurobiology, psychology, and cultural anthropology of olfaction and taste, respectively.

CHAPTER 5

THE VISUAL SYSTEM

KEY THEMES

- The anatomical connectivity of the visual system creates a substrate for two broad classes of visual information processing: determining *what* an object is, and determining *where* it is.
- Carefully designed lesion studies have localized these two categories of function to the temporal and parietal cortex, respectively.
- Within this organizing principle, there is considerable cross-talk and integration of information.
- Visual object recognition is accomplished via a progressive hierarchy of processing steps, beginning in V1 and extending into inferotemporal (IT) cortex.
- Although one logical terminus of such a hierarchy might be an apex level of extreme localization, with one neuron representing each object that we've encountered in our lives, the data suggest a more plausible distributed combination-coding scheme.
- Since the 1960s, the processing of faces in the temporal cortex has been a focus of interest, and a source of puzzlement, because faces seem to follow different "rules" than other types of complex object.
- Some populations of IT neurons seem to encode "the idea of a face," rather than the identity of any one specific face, thereby suggesting an explicit interface between perception and memory.
- Visual illusions offer an effective tool for probing the functions of feedback in the visual system.
- An important role for feedback is to disambiguate conflicting bottom-up signals from early stages of the visual system.

CONTENTS

KEY THEMES

FAMILIAR PRINCIPLES AND PROCESSES, APPLIED TO HIGHER-LEVEL REPRESENTATIONS

TWO PARALLEL PATHWAYS

A diversity of projections from V1

A functional dissociation of visual perception of
what an object is vs. *where* it is located

Interconnectedness within and between the two
pathways

THE ORGANIZATION AND FUNCTIONS OF THE VENTRAL VISUAL PROCESSING STREAM

Hand cells, face cells, and grandmother cells

*Electrophysiological recordings in temporal
cortex*

*Broader implications of visual properties of
temporal cortex neurons*

A hierarchy of stimulus representation

*Combination coding, an alternative to the
“grandmother cell”*

A critical role for feedback in the ventral visual
processing stream

*Visual illusions provide evidence for top-down
influences*

*Top-down feedback can disambiguate the
visual scene*

TAKING STOCK

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

FAMILIAR PRINCIPLES AND PROCESSES, APPLIED TO HIGHER-LEVEL REPRESENTATIONS

This chapter will address the question of how we recognize objects that we encounter in the world. In one sense it will represent our first foray into the unequivocally cognitive, in that recognizing what an object is necessarily entails *thinking* about where you've seen it before and what its significance is for you (e.g., *I see my [spouse/friend/professor/etc.]*), or, if it's an exemplar that you've never seen before, *thinking* about how it relates to what you already know about the world (e.g., *It's the face of a man/woman/young person/old person*) and what an appropriate action when encountering such an object might be. And so, as this preamble implies (as, indeed, does the word *recognition*), this is a domain that necessarily engages memory – stored representations of knowledge about the world; the very stuff of thought.

The emphasis of this chapter will be on the sensory modality of vision, with consideration given to how information from multiple sensory modalities can interact to influence how we perceive the world. At a detailed, mechanistic level, we shall see that a principle that we examined in *Chapter 3* also underlies object recognition. In that earlier chapter we saw that features such as orientation specificity, binocular disparity, and direction selectivity are constructed by neurons in V1 via selective sampling across inputs from LGN neurons that have much simpler receptive field properties. At a simplified level of description, the same principle, playing out in parallel in hundreds of thousands of neurons, is believed to underlie object recognition – just the right set of features, located in just the right part of the visual field, combine to represent a complex object. Thus, there is necessarily a hierarchical element to object recognition, in that many simpler elements are combined at lower levels to construct progressively more complex representations at higher levels. This idea of many feeding into one will lead us to revisit the tension introduced in *Chapter 1*, between localization vs. mass action (reframed here as localist vs. distributed representation). We shall also see that the principle of feedback from higher to lower

levels of representation, first considered in detail in *Chapter 3*, is also characteristic of the hierarchies involved in recognizing objects. This will also invoke the principle of predictive coding, the idea that a neural representation of the world is already present in the brain, and that the brain uses incoming sensory inputs to update this representation.

As we have done previously, we'll start by considering the anatomical structure of the systems that support high-level vision.

TWO PARALLEL PATHWAYS

A diversity of projections from V1

You won't be surprised to learn that *Chapter 3*'s treatment of area V1, and of that region's interactions with MT, glossed over some of the intricacies both of V1's functional organization, and of its feedforward projections. As is illustrated in *Figure 5.1*, there is both a direct route from V1 to MT, which carries signals primarily from the magnocellular pathway, and an indirect route, with neurons from V1 first projecting to V2, and V2 neurons projecting to MT. The indirect pathway carries information from both the magnocellular and parvocellular pathway, a result of the convergence in layer IV of V1 of inputs from magno- and parvocellular pathways on the neurons projecting to V2. V2, like V1, also shows heterogeneous cytochrome oxidase staining, with regions appearing as thick stripes, thin stripes, or interstripes (*Figure 5.1*). The indirect pathway from V1 to MT synapses in the thick stripe region of V2. MT, in turn, sends feedforward projections to area MST (located in *medial superior temporal* cortex in the macaque brain), which, in turn, sends feedforward projections to several areas in the posterior parietal cortex (PPC).

Separate from the direct and indirect projections to MT, different populations of V1 neurons project to the thin-stripe and interstripe regions of V2. These regions of V2, in turn, send feedforward projections to area V4 of the temporal cortex, with V4's feedforward projections extending into the IT cortex (See *Tip 5.1*). More detail about the anatomical connectivity of the visual system appears further along in this section, but at this point we'll focus on the fact that there are two

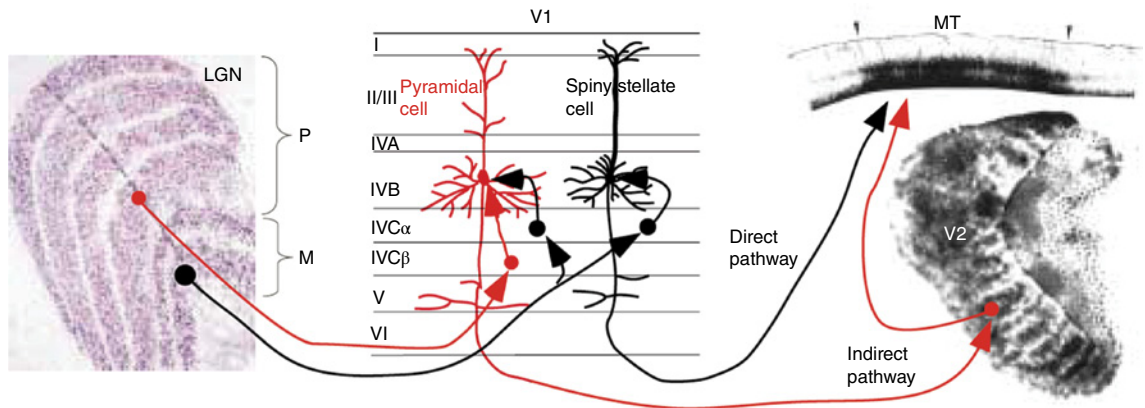


FIGURE 5.1 Dual routes from LGN to MT. Left-hand panel illustrates a cross section through the LGN, highlighting the two layers containing magnocellular (M) neurons, and the four layers containing parvocellular (P) neurons. Superimposed in black ink is a cartooned magnocellular neuron projecting to two neurons in layer IV α of cortical area V1. Superimposed in red ink is a cartooned parvocellular neuron projecting to a neuron in layer IV β of cortical area V1. Of the two IV α neurons receiving magnocellular input, one projects to a “spiny stellate” neuron in layer IV β , which, in turn, projects directly to area MT, the other to a pyramidal neuron in layer IV β . The layer IV β neuron also projects to this layer IV β pyramidal neuron, which can thus be characterized as “blending” input from both the M and P channels. The IV β pyramidal neuron projects to a neuron in the thick stripe region of cortical area V2 which, in turn, also projects to MT. This figure shows a portion of the surface area of V2 that has been stained for cytochrome oxidase (*Methodology Box 3.1*); note that immediately to the right of the thick stripe highlighted in this image is an interstripe and, to the right of the interstripe, a thin stripe. Also note that, in this image, the portion of V2 that is illustrated is shaped like a boomerang, with a clear discontinuity between it and the cortical tissue bordering it to the right. The region to the right is V1, identifiable from the small spots (i.e., *blobs*) of cytochrome oxidase staining. Finally, MT is illustrated here in cross section, with deep layers at the bottom of the image, and has been stained to emphasize one of its distinguishing characteristics, which is heavy myelination of fibers even after they have entered the cortex. (Note that this dark staining ends abruptly near the middle of the cortex; why should that be?). Source: Born, 2001. Reproduced with permission of Elsevier.

Tip 5.1

Convention has it that posterior parietal cortex is abbreviated PPC, whereas inferotemporal cortex is abbreviated IT (or, IT cortex).

pathways originating in V1, one that projects through the thick stripes of V2 to MT, MST, and PPC, a second that projects through the thin- and interstripe regions of V2 to V4, then IT. As suggested by this anatomy, there are important, and importantly dissociable, visual functions supported by the posterior parietal and temporal lobes.

Insights into the visual functions of the PPC and IT have come from both of the scientific traditions

highlighted at the end of *Chapter 1*: behavioral neurology/neuropsychology; and systems neuroscience. Those clearly associated with the former include case studies detailing visual impairments in patients whose strokes have damaged circumscribed regions of cortex. Those clearly associated with the latter include extracellular electrophysiological studies in monkeys performing visual discrimination tasks, and detailed studies of neuroanatomy. Many examples of this research will be considered further along in this chapter, but we will start examining these functions via a study performed in the late 1960s at a laboratory firmly grounded in both scientific traditions: the Laboratory of Neuropsychology at the National Institutes of Health.

A functional dissociation of visual perception of what an object is vs. where it is located

If one were forced to choose a label for the Section of Neuropsychology (as it was then called) in the 1960s and 1970s, one would classify it as a systems neuroscience laboratory – the primary experimental approach was one of testing hypotheses about the neural bases of cognitive functions by surgically lesioning specific brain regions and assessing the effects on carefully designed behavioral tasks. Many of those in the lab, however, also had formal training in human neuropsychology, studying behavioral abnormalities in neurological patients. In this environment, a graduate student named Walter Pohl conducted a series of experiments that, in retrospect, seem to have been “made to order” for this textbook writer. I say this for three reasons: first, they illustrate an important principle of the functional organization of the visual system; second, they illustrate a key principle of inference in neuropsychological research (to be detailed in *Methodology Box 5.1*); and third, they illustrate the value of performing thorough, carefully controlled experiments, in that an experiment that was intended at the time as a “mere” control condition turned out to provide critical, substantive information.

The primary question of interest for Pohl was different from the one that we are currently addressing. It was motivated by the idea that “The spatial position of an object may be defined in either of two ways, [...] either by reference to the position of the observer himself who serves as the spatial anchor, or by reference to the position of another object which functions as an external landmark” (Pohl, 1973, p. 227). In particular, a previous study of brain-injured war veterans conducted by neuroscientist Josephine Semmes and colleagues had shown that whereas some had difficulty viewing a mark on a picture of a body and pointing to the corresponding location on their own body, others had difficulty following maps. The former is referred to as **egocentric localization** (i.e., using a *self*-centered frame of reference), the latter as **allocentric localization** (i.e., using an *other*-centered frame of reference). Thus, Pohl’s experiments weren’t intended to be about the visual functions of posterior parietal and IT cortex, per se, but rather about two ways of perceiving space,

and of acting on objects in the environment. (The topic of spatial frames of reference will be revisited greater detail in *Chapters 6–8*, on spatial cognition and skeleto- and oculomotor control, respectively.) It is the inclusion of a critical control experiment that makes them so relevant for our present purposes.

As illustrated in *Figure 5.2*, and detailed in *Research Spotlight 5.1*, Pohl (1973) collected data from monkeys performing a “landmark discrimination” task, intended to assess the ability to judge the location of a “landmark” stimulus, and an “object discrimination” task, intended to assess visual object recognition. Depending on the experiment (four are reported in Pohl (1973), but only two are emphasized here), the monkeys had surgical lesions of PPC, IT, prefrontal cortex, or were unoperated control subjects. The results from these experiments showed persuasively that PPC is selectively important for performance on the landmark task, but not for performance on the object discrimination task. Conversely, IT cortex was shown to be selectively important for the object discrimination task, but not the landmark task. The particular pattern of these results – that damage to area A affected performance on task X, but not on task Y; whereas damage to area B affected performance on task Y, but not on task X – is important, because of its implications for the specificity of the functions supported by these two brain areas. These ideas were first explored in *Methodology Box 1.1*, and now are further elaborated on in *Methodology Box 5.1*, and codified as the **double dissociation of function**. In interpreting his findings, Pohl (1973) also drew on many previous theories and findings to advance the interpretation that regions of PPC may be “specialized for extrafoveal visual functions, as in the perception of spatial relations among objects in the visual field [... whereas IT cortex may be] specialized for perceptual functions related to foveal vision, as in object and pattern discrimination” (p. 238). (*Figure 5.3* illustrates what is meant by extrafoveal vs. foveal vision. See *Thought Box 5.1* for a consideration of the lesion method.)

The profound importance of the Pohl (1973) study was manifest a few years later when Pohl’s mentor at the Laboratory of Neuropsychology, Mortimer Mishkin, together with colleague Leslie Ungerleider, published a thorough review of what was known about the anatomical connections of occipital-to-temporal

RESEARCH SPOTLIGHT

5.1 What vs. where in the primate visual system (Pohl, 1973)

The behavioral tasks, administered in a WGTA (*Figure 5.2.A*), were variants of the reversal learning procedure, which accommodates the fact that, unlike human research subjects, one cannot give animals verbal instructions. Instead, the rate at which an animal can learn the rules for the task by trial-and-error guessing is used as one of the primary dependent measures. In the landmark reversal task, for example, *Figure 5.2.D* illustrates what the monkey might see on the very first trial of the experiment. The food reward was always hidden in the well adjacent to the landmark stimulus. If the monkey reached correctly, it retrieved the reward, the screen was lowered, and the tray prepared for the subsequent trial. If the monkey incorrectly reached for the well that was distant from the landmark, it would find this well to be empty. On such error trials, feedback was given on a subsequent correction trial by re-presenting to the animal the same stimulus array, but with the incorrect food well now uncovered, so that the animal could see that it contained no reward. Regardless of whether the prior trial had been performed correctly or incorrectly, on the subsequent trial there was an equal likelihood that the landmark would appear on the right or on the left.

Each day the monkey performed 30 trials in this fashion (not counting correction trials) until the experimenters deemed it to have learned the rule, defined as getting at least 28 of 30 trials correct on one day's testing (i.e., ~ 93%), followed by a subsequent day of getting at least 24 of 30 trials correct (i.e., ~ 80%). The primary measure of performance was the number of errors that monkeys made before reaching this level of performance. Note that ending the experiment here, however, would leave interpretation of the results ambiguous, for two reasons. One is that this criterion for learning, although reasonable, seems arbitrary. Second, what if, for example, a given monkey was just "slow to catch on" to the rule, independent of the cognitive function that the task is intended to capture? (Note from *Figure 5.2.H*, for example, that the IT-lesioned group performed worse than the PPC-lesioned group on the first set of trials, as, indeed, did the unoperated control animals!) As *Figure 5.2.H* and *Figure 5.2.I* illustrate, however, including the reversal element to the experimental design largely alleviates such concerns about subjectivity and experimental error: on the day after the monkey had achieved criterion-level performance, the rule was reversed. That is, from this day forward, the reward would be placed in the food well that was not adjacent to the landmark. As on day one, the monkey had to learn the new rule by process of trial-and-error. In this way, the experiment proceeded, through a total of seven reversals.

The object reversal task (*Figure 5.2.E*) followed the same procedure, with the exception that the animals had to learn, on each set of trials, under which of two distinctive objects (red striped cylinder vs. a silver cube) the reward was located. The results from this study, and how they were interpreted, are discussed in *Methodology Box 5.1*.

METHODOLOGY BOX

5.1 The double dissociation of function

Pohl (1973) performed the two studies illustrated in *Figure 5.2.D* and *Figure 5.2.E* in that order – first the landmark task, then the object discrimination task. Focusing for a moment on the results of just the former (*Figure 5.2.H*), one can see that this, alone, is a difficult pattern of results to assess. First, as noted in *Research Spotlight 5.1*, all groups had difficulty on the first three sets of trials. Could it not be the case, therefore, that damage to parietal cortex simply makes it difficult to perform well on difficult tasks? (E.g., what if the true effect of the PPC lesions were to produce an incessant, loud ringing in the ears, such that the animals would have difficulty holding their concentration on any difficult task for a long period of time? If this were true, the PPC-lesioned animals would also have performed worse on, say, a difficult olfactory discrimination task.) A more subtle concern was raised by Pohl to motivate the object discrimination task: because the IT animals showed exaggerated difficulty on first few sets of trials, might it not be that both the IT lesions and the PPC lesions impaired the ability to perform an object discrimination? (E.g., what if the true explanation for the results was that animals in both groups had difficulties discriminating a simple plaque covering a food well from a plaque with a cylinder touching it?) Because neither of these alternative accounts could be ruled out with the data from the landmark reversal task alone, an additional experiment was needed.

For this additional experiment, Pohl (1973) selected a task that he suspected would depend on the functioning of IT cortex, a visual discrimination task. Indeed, as illustrated in *Figure 5.2.I*, results on this task showed the opposite pattern from the landmark reversal task: object discrimination reversal was selectively impaired in the IT-lesioned animals. Critically, the PPC-lesioned animals performed the task almost flawlessly. Consideration of these two sets of results together permits one to rule out the two alternative explanations for the results from the landmark reversal task. With regard to the difficulty explanation, if a PPC lesion produces a quantitatively larger deficit on any type of task, then it would have done so on the object discrimination task. Because it didn't, this alternative explanation can be ruled out. With regard to the underlying nature of the impairment of the PPC-lesioned group on the landmark discrimination task, if this had been an object-discrimination deficit, then the PPC-lesioned group should also have been impaired on the task that clearly made stronger demands on object discrimination. Because it wasn't, however, this alternative can also be ruled out.

The logic laid out in the previous paragraph permitted Pohl (1973) to reason that “the lesion effects [in these two studies] differ in kind” (p. 234). That is, the double dissociation produced across these two experiments supports the conclusion that lesions of these two brain regions produce behavioral deficits that differ qualitatively, and thus suggests that different functions are supported by these two systems. (Note that, for those familiar with the analysis of variance (ANOVA), a double dissociation is analogous to a crossover interaction.)

FIGURE 5.2 The experimental procedures used by Pohl (1973).

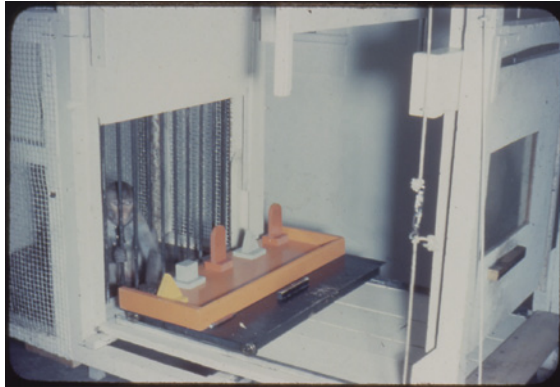


FIGURE 5.2.A Photograph of a macaque monkey performing a task in a Wisconsin General Testing Apparatus (WGTA). Note that the animal is in a cage, and can interact with items on the wooden testing tray by reaching through the bars in the cage. The tray is on wheels and can be pulled away or toward the animal via ropes on pulleys. A screen can be lowered such that the animal can neither see, nor reach, through the bars of the cage while a trial is being prepared. On the right side of the image is a one-way mirror, through which an experimenter can observe the animal's behavior, and from where s/he can control screen and tray. Source: Harlow Primate Laboratory, University of Wisconsin, Madison.

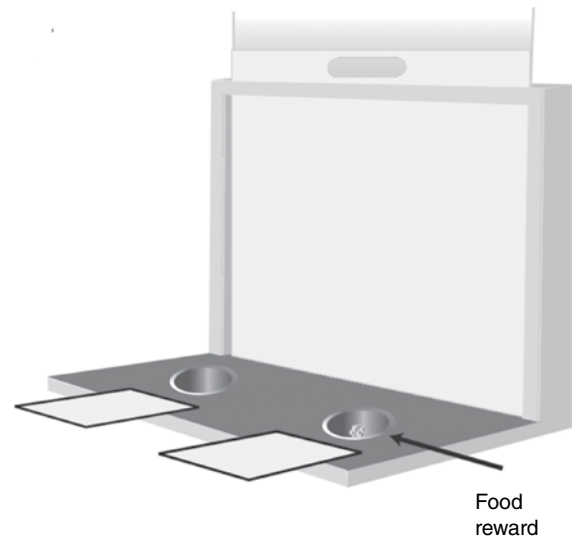


FIGURE 5.2.B Illustration of a two-food-well configuration of the testing tray, with a food reward being placed in one of the two wells (Pohl (1973) used a half peanut on each trial). Source: Adapted from Curtis and D'Esposito, 2004.



FIGURE 5.2.C Illustration of a monkey reaching for one of the two covered food wells. Because the animal cannot see the contents of either food well, it must make its choice based on some second-order factor not related to the wells themselves (e.g., a guess, a memory of the previous trial (these would be considered “internal” factors, because they originate with the animal itself), or, as illustrated in panels *D* and *E*, the configuration of stimuli in relation to the two wells.) Source: Adapted from Curtis and D'Esposito, 2004.

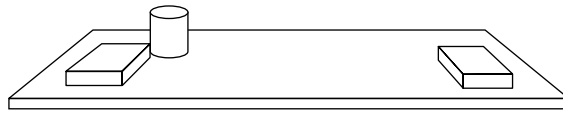


FIGURE 5.2.D The monkey's view of the testing tray during a trial of the landmark reversal (i.e., spatial discrimination) task. The landmark was a red cylinder with vertical white stripes. Source: Pohl, 1973. Reproduced with permission of APA.

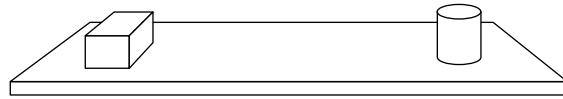


FIGURE 5.2.E The monkey's view of the testing tray during a trial of the object reversal (i.e., object discrimination) task. Each object is placed directly on top of a food well. Source: Pohl, 1973. Reproduced with permission of APA.

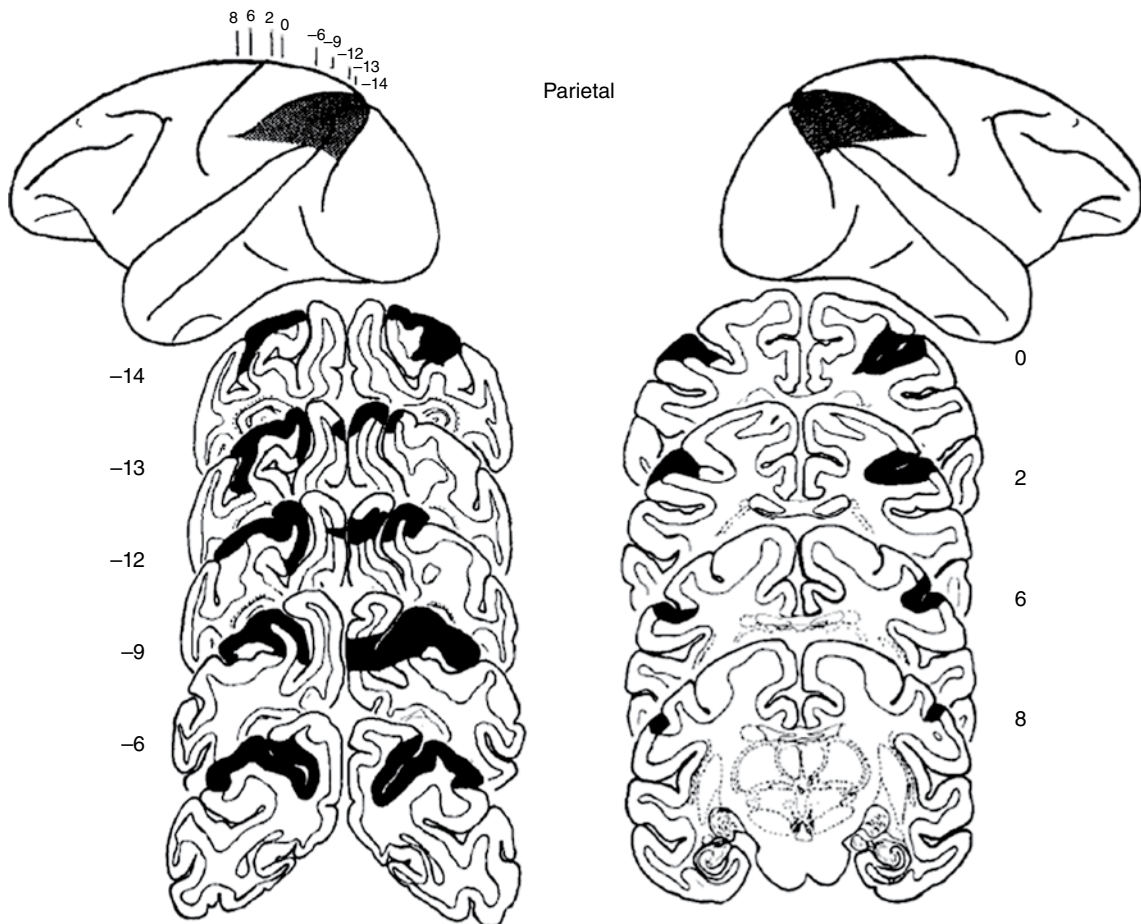


FIGURE 5.2.F Illustration of a PPC lesion. The numbers in the lateral view of the right hemisphere correspond to each of the coronal sections shown below, and also refer to the distance, in millimeters, from an origin point in an atlas of the monkey brain. Note that although the figure is labeled “parietal,” great care was taken that the lesions did not invade more anterior areas known to be important for somatosensation (as detailed in *Chapter 4*). Source: Pohl, 1973. Reproduced with permission of APA.

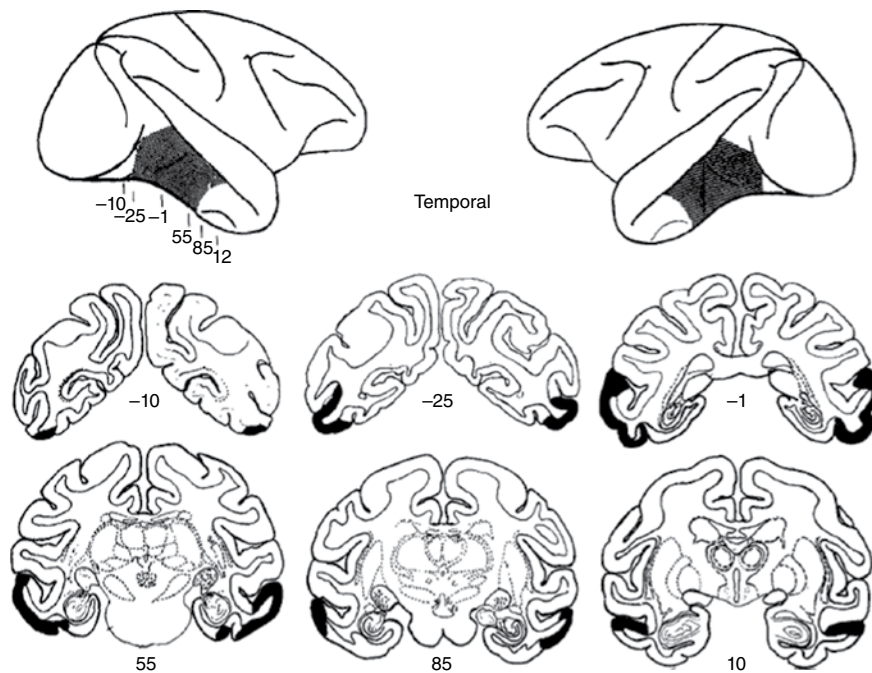


FIGURE 5.2.G Illustration of an IT lesion, same conventions as *Figure 5.2.F*. Source: Pohl, 1973. Reproduced with permission of APA.

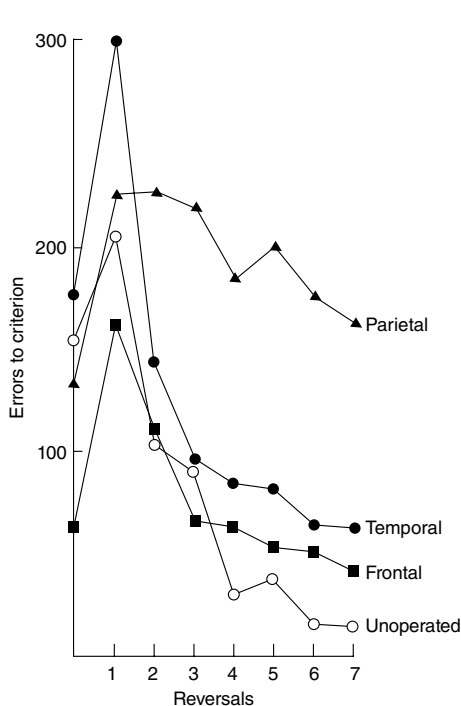


FIGURE 5.2.H Behavioral performance of monkeys from four experimental groups on the landmark reversal (i.e., spatial discrimination) task, illustrating markedly inferior performance by the parietal-lesioned animals. Source: Pohl, 1973. Reproduced with permission of APA.

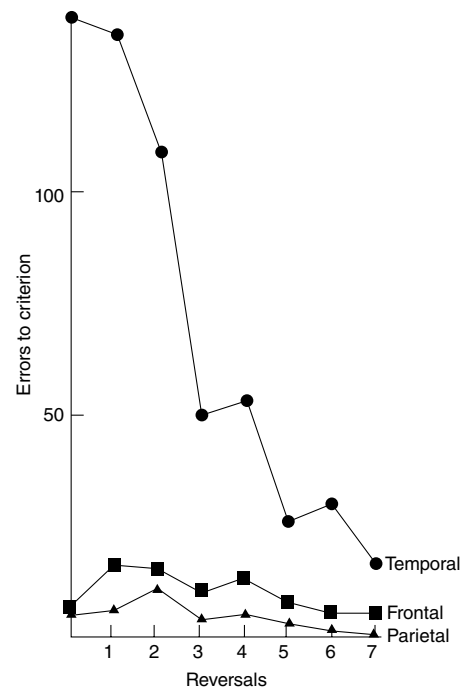


FIGURE 5.2.I Behavioral performance of monkeys from three experimental groups on the object reversal (i.e., object discrimination) task, illustrating markedly inferior performance by the inferior temporal-lesioned animals. (An unoperated control group was not included in the object reversal testing.) Source: Pohl, 1973. Reproduced with permission of APA.

THOUGHT BOX

5.1 The kind, and strength, of inference supported by different experimental methods

The following question might have occurred to some while reading this section: *The Pohl (1973) study was a lesion study, using experimental methods not dramatically different from what Ferrier and Munk had used 100 years previously; doesn't this make it's methods "old fashioned" and, therefore, perhaps suspect, or somehow less valid?* The answer to this is an emphatic "No," according to at least two lines of argumentation. The first is that, when one considers the details of the experimental procedure – the theory guiding task design; the precision and systematicity of how behavioral measures were defined and how performance was scored; the control afforded by the WGTA, and related minimization of experimental confounds; the precision of the surgical lesions; the formal logic of the double dissociation – the Pohl (1973) study was markedly superior, from a technical standpoint, to those from the century before. The second line of argument is that, for the same reason that "correlation does not imply causation," it remains just as true today as it did in 1973, and as it did in 1873, that one cannot infer the necessity of a system's contribution to a function without perturbing the system and observing that this perturbation produces systematic alterations in the function.

Depending on the specific question being asked, there are many different ways to produce an experimental perturbation. The damage or removal of tissue is the method that has received the most attention thus far in this text. Falling under this category are surgical lesions, produced experimentally in animals or for clinical purposes in human patients, and so-called "accidents of nature" – be they head injuries, acute neurological trauma (e.g., stroke, **anoxia**, **viral encephalitis**), or neurodegenerative disorder. A second perturbation technique is stimulation. Although stimulation can be used to activate a system, as when electrical microstimulation of the motor cortex produces the movement of a limb (e.g., Fritsch and Hitzig (1870); *Chapter 1*), it can be delivered in a way that disrupts the normal functioning of a system. For example, *Chapter 9* will consider an experiment in which perturbation delivered via electrical microstimulation has provided strong evidence in relation to a problem that has long vexed the cognitive neuroscience of visual perception, which is whether the processing of information about faces recruits specialized regions of IT cortex. In subsequent chapters we will see how transcranial magnetic stimulation (TMS) can be used to perturb normal brain function. Finally, pharmacological lesions can be produced by delivering chemical agents that can, for example, block a particular class of neurotransmitter receptor or a particular type of membrane-spanning ion channel.

Of course, no single class of experimental method can address all the questions that can be asked about how the brain gives rise to the mind. In particular, there are many questions of mechanism, such as *How is object recognition accomplished by the visual system?*, that the lesion method, alone, is ill-suited to address. For questions such as this, a way to measure the activity of a neuron, or a network, or a region of tissue, and how this activity varies with experimental manipulations, can be necessary.

and occipital-to-parietal cortex, about the effects of lesions to IT vs. PPC, and about the electrophysiological properties of neurons in IT vs. PPC (Figure 5.3), and argued that the former constituted a system for analyzing *what* an object is, and the latter a system for analyzing *where* an object is (Figure 5.4). This model

has been highly influential in cognitive neuroscience, to the point that it has acquired two sets of nicknames, a functional one – *what* vs. *where* – and an anatomical one – *ventral stream* vs. *dorsal stream*.

Interconnectedness within and between the two pathways

As is often the case with conceptual models of brain function, the details of the functional organization of the visual system turn out to be much more complicated than what is suggested, for example, by Figure 5.4. One example relates to information carried by the M vs. the P channels from the retina through V1, and onto higher levels of the visual system. The two-parallel-pathways model seems to fit quite well with the differential projection patterns, considered earlier in this section, of neurons in thick- vs. thin- and interstripe regions of area V2. Indeed, this led some to suggest that the dorsal visual processing stream may carry primarily magnocellular signals, whereas the ventral stream may carry primarily parvocellular signals. This logic led to a research strategy that sought to exploit this presumed correspondence of the M and P channels with the dorsal and ventral pathways, respectively, by presenting stimuli under **isoluminant** conditions. That is, the stimulus display was controlled such that all objects, including the background, were presented with the same brightness (the same *luminance*), and so only differences in chrominance (i.e., color) could be used to discriminate objects from each other, and from the background. The logic was that an isoluminant display would require exclusive reliance on the ventral stream, because the dorsal stream was presumed to be dominated by the M channel. Isoluminance, therefore, might be a way to study the functions of the ventral stream while experimentally keeping the dorsal stream “offline.”

Follow-up studies, however, have called the logic of testing at isoluminance into question. One persuasive line of work, by Peter Schiller and his colleagues, has examined the effects of selectively lesioning one channel or the other (via small lesions in selective layers of the LGN), then testing performance on tasks that require visual processing characteristic of the dorsal stream or the ventral stream. These studies did, indeed, find preferential impairment on

FIGURE 5.3 Electrophysiological properties of neurons characteristic of the *what* and *where* systems. For both panels, the vertical and horizontal axes represent the meridia of the visual field (“+” indicating upper or right, and each tick mark indicating 10° of visual angle). Thus, the intersection of the two axes corresponds to the fovea. The shape to the left of each panel represents a slow-moving stimulus that would typically be effective at eliciting a response from a neuron in the region in question.

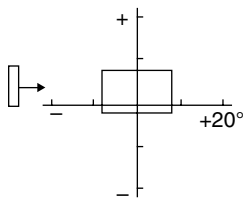


FIGURE 5.3.A For IT neurons, 60% have receptive fields that span the vertical midline (i.e., they are bilateral), and the receptive fields of 100% of them include the fovea. These neurons are typically selective for particular stimulus features. Source: Ungerleider and Mishkin, 1982. Reproduced with permission of The MIT Press.

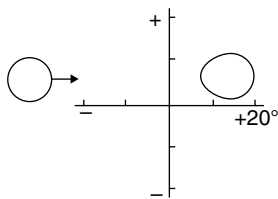


FIGURE 5.3.B For PPC neurons, the receptive fields of > 60% of them are restricted to the contralateral visual field, and the receptive fields of > 60% of them do NOT include the fovea. At the time it was believed that these neurons did not show selectivity for stimulus features (a perspective that has since been disproven for at least some regions of the posterior parietal cortex).

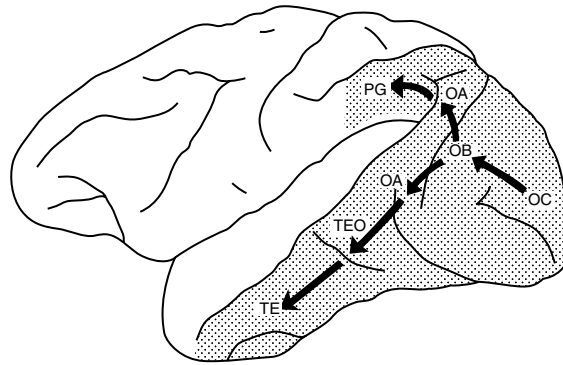


FIGURE 5.4 Schematization of the *what* and *where* pathways of the dorsal and ventral visual processing streams in the macaque monkey. The labels correspond to names given these regions in a neuroanatomy atlas: OC is striate cortex; OB and OA are extrastriate occipital areas; PG is posterior and inferior parietal cortex; TEO and TE are IT cortex. Note that the sulcus forming the dorsal boundary of area PG is the intraparietal sulcus, which will feature prominently in chapters on attention, skeleto- and oculomotor control, and decision making. Source: Mishkin, Ungerleider, and Macko, 1983. Reproduced with permission of Elsevier.

“ventral-stream tasks” after lesions of the P system, and of “dorsal-stream tasks” after lesions of the M system, but the impairments were relatively subtle, not the devastating effects that one would predict from frank lesions of the temporal or parietal cortex. Thus, for example, parvocellular lesions impaired discrimination of textures with high spatial frequencies, but not those with low spatial frequencies. Magnocellular lesions, in contrast, impaired the ability to discriminate a flickering light from a steady light at high, but not low, rates of flicker. Similarly, they impaired the ability to detect a difference in the speed of motion of a small group of dots within a larger field of moving dots, but only when the difference in speed was small, and thus difficult to detect. In both cases, deficits were exacerbated when the to-be-detected stimulus was presented at low contrast (e.g., the impairment was worse for a dim light flickering against a gray background than for a bright light). Schiller and colleagues concluded that the important contribution of the M and P channels is to extend the range of performance of the visual system in two dimensions – spatial frequency and temporal – under challenging conditions. Thus, the P channel extends the visual system’s resolution for fine-grained discriminations of texture and color, and the M channel extends its temporal resolution.

One implication of the findings of Schiller and colleagues was that the M and P pathways each contribute to functions of both the dorsal and the ventral streams (but see *Tip 5.2*). Consistent with this is what we already saw in *Figure 5.1*, that the anatomy indicates that there is already considerable blending of inputs from magno- and parvocellular neurons at the level of V1. Additionally, as portrayed in *Figure 5.5*, we now know that there are many instances of direct, reciprocal interconnections between regions of the two streams considered to be at comparable levels of the visual processing hierarchy, such as between V4 of the ventral stream and MT of the dorsal stream.

Tip 5.2

An exception to this statement is the perception of color, which is heavily dependent on the P system. Nonetheless, although P-system lesions abolish the capacity to identify or discriminate among colors, they do not abolish the ability to perceive objects that are defined only by color (i.e., that are isoluminant with the background display).

FIGURE 5.5 Anatomical connectivity of the visual system. See Felleman and Van Essen, 1991. Source: Felleman and Van Essen, 1991. Reproduced with permission of Oxford University Press.

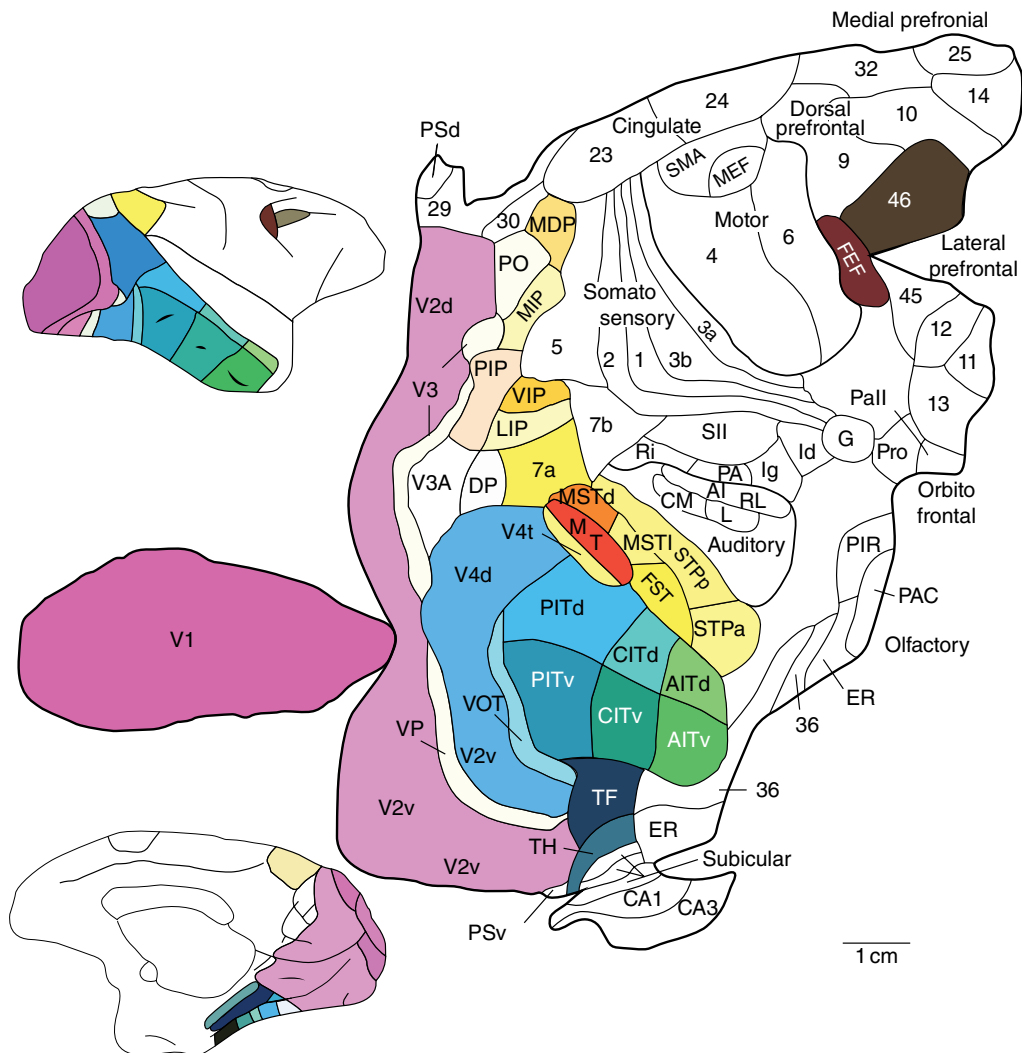


FIGURE 5.5.A Color-coded views of the right hemisphere of the macaque cerebral cortex (plus hippocampus) as seen from lateral (top-left inset) and medial (bottom-left inset) views, and as a virtual “flat map” that illustrates the entire cortical surface. To create a flat map that preserves the relative size of each region (e.g., that doesn’t “stretch” or otherwise distort), it is necessary to make several virtual “cuts.” Most obvious in this view is that V1 has been “cut out” along the entirety of its border with V2, with the exception of its anterior-most point on the lateral surface. The next major cut is on the medial surface, “carving” around the roughly circular boundary of cortex and non-cortex (revisiting *Figure 1.2.B* and *Figure 2.2* may be helpful here). At this point, one can imagine approaching the brain from the lateral view and reaching one hand, palm down, over the parietal cortex and down along the midline, the other hand, palm up, under the ventral surface of the temporal lobe with fingers curling up and in along its medial surface, and then pulling up with the top hand and down with the bottom so as to “unroll” this 3-D shape and flatten it. If, in the process, one were also to gently stretch by pulling the two hands further apart, the interior of the sulci would also be exposed. It is thus, for example, that the red and yellow band of tissue that is not visible in the lateral view, by virtue of lying within the superior temporal sulcus, is exposed in the flat-map view. Finally, two additional cuts are needed to complete the task, each separating regions that are adjacent in the intact brain. One is a cut separating prefrontal areas 46 from 45 (running approximately along the fundus of the principal sulcus), the second is a slit starting at the temporal pole and proceeding caudally along the medial surface, thereby artificially dividing the entorhinal cortex (ER) in two. This type of flat map is also often used to display results from fMRI studies of the human brain, as we shall see further along in this chapter.

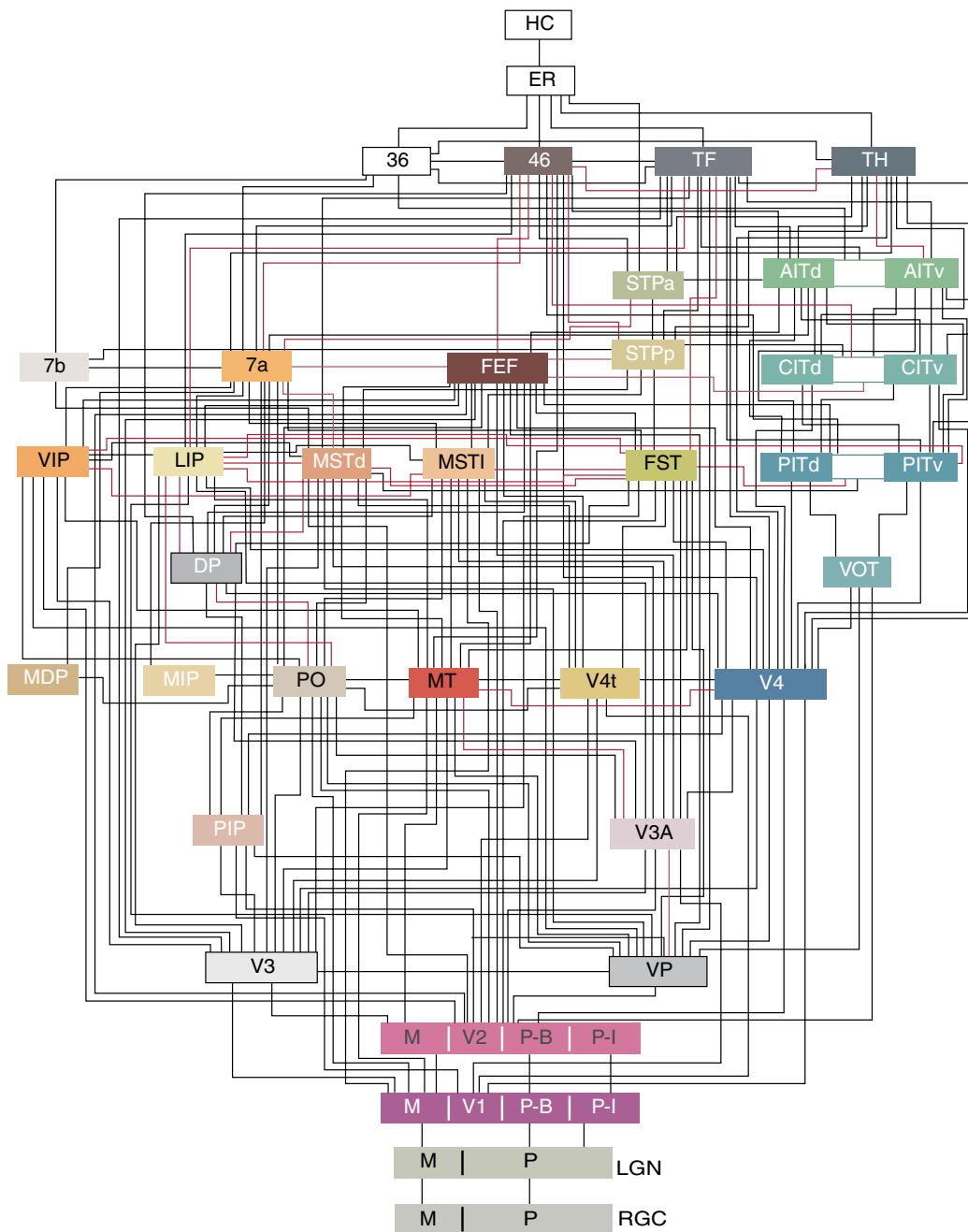


FIGURE 5.5.B Connectivity diagram of the visual system as a hierarchically organized set of interconnected regions, with each region's position in the hierarchy determined by its anatomical connectivity with all the others, as determined by laminar patterns of its inputs and outputs (see *Figure 2.4*, and associated text from *Chapter 2*). Color-coding corresponds to *Figure 5.5.A*. Each line is understood to correspond to bidirectional connections between two connected regions, with feedforward connections going from lower to higher on the diagram (e.g., from MT to one of the MST subregions) and feedback connections the opposite. Thus, each horizontal level of the diagram corresponds to a putative "hierarchical level of cortical processing." The bottom level is the retina, with retinal ganglion cells (RGC) divided into M and P channels, which project to the magnocellular (M) or parvocellular (P) layers of the LGN (second level). The first two levels of cortex, V1 (purple) and V2 (pink), are each divided into sectors that carry unequivocally M or P-blob (P-B) or P-interblob (P-I) signals; connections from regions that blend these signals are labeled "V1" and "V2."

Two final comments about the *what* vs. *where* model before we delve into the functions of the ventral stream. First, it's noteworthy that *Figure 5.4* omits the feedback projections that we already know from earlier chapters make fundamental contributions to how the brain works. This will be taken up near the end of this chapter. Second, an additional level of complexity that we won't consider in detail until *Chapter 7* (skeletal motor control), harkens back to *Thought Box 1.1*'s question of *what's a function*: some argue that the dorsal stream is better construed as carrying out the function of *how* the body carries out actions in space, rather than of determining *where* objects are located in space (*Tip 5.3*).

Having now established some gross anatomical boundaries for the visual system, and a framework for thinking about how the functions of this system are organized (i.e. *what* vs. *where*), we can proceed to consider the question *How does visual object recognition work?* In the process, some of the ideas left underspecified in the preceding paragraph, such as that of the “visual processing hierarchy,” will be fleshed out. For the remainder of this chapter, periodically returning to *Figure 5.5* will be helpful, and, in the process, its “subway system wiring diagram” will become progressively more familiar and less daunting!

Tip 5.3

This highlights a reality about which scientists must always be mindful: just because one designed an experiment with a particular function in mind (in this case, spatial perception) does not necessitate that performance on the task engages this function. (E.g., it's possible that Pohl's PPC-lesioned monkeys displayed impairment of a *how* system, rather than of a *where* system, despite the intentions of the experimenters.) Often, the hypothesized function is just that – a theoretical construct – and it is important not to conflate a theoretical construct with data produced by a particular task.

THE ORGANIZATION AND FUNCTIONS OF THE VENTRAL VISUAL PROCESSING STREAM

Hand cells, face cells, and grandmother cells

Electrophysiological recordings in temporal cortex

By the 1960s, decades of clinical observations in humans and lesion studies in experimental animals had established that the temporal lobes play a critical role in object recognition. What remained unknown, however, was *how* this region carried out that function. *What were the mechanisms that gave rise to object recognition (Thought Box 5.1)?* It was the pioneering studies of Hubel, Wiesel, and their colleagues, of the functions of V1 (*Chapter 3*), that provided the methodological and conceptual framework that could be applied to this question. Among the first to do so were Charles Gross and his colleagues, who began recording from neurons in IT cortex, and subsequently the superior temporal sulcus, in the monkey in the 1960s and 70s. Because the previous literature indicated that lesions of IT cortex produced profound object recognition deficits, they began by measuring responses to a wide array of complex objects.

In a narrative reminiscent of that from Hubel and Wiesel's early experiments, Gross and colleagues initially had difficulty finding any type of object that would reliably elicit a response. Their first successful recording came when, after failing to activate a neuron with a large set of stimuli, an experimenter waved his hand in front of the display screen, and the neuron responded to it vigorously. *Figure 5.6* illustrates a series of stimuli that the researchers subsequently presented to this neuron. There is a subjective sense from this stimulus set that the closer the stimulus was to resembling a monkey's hand, the more effective it was at activating the neuron. Thus, it didn't seem to be the case that this neuron's preferred stimulus was a more “primitive,” or generic shape, and that the hand just happened to have many of the same characteristics of this neuron's preferred stimulus characteristics. Rather, it seemed that this neuron's preferred stimulus was truly the hand, and other stimuli could “drive” it with

more or less effectiveness as a function of how closely they resembled the hand (*Tip 5.4*). Another noteworthy property of this neuron was that it responded to the hand stimulus with a comparably vigorous response when it was positioned in any of many locations in the visual field. That is, it had a very large receptive field. Although this experiment did not find other neurons with the same selectivity for the hand stimulus shown in *Figure 5.6*, it did find that many visually responsive neurons in IT that responded better to “complex” stimuli than to simple geometric shapes. Typically, each of these neurons had its own unique preferred stimulus. The property that IT neurons had in common, however, was large receptive fields – often larger than $10^\circ \times 10^\circ$ of visual angle – within which the neuron would respond equally well to its preferred stimulus. (This property was highlighted by Ungerleider and Mishkin (1982), and is illustrated in *Figure 5.3*) One way of describing this property of an IT neuron is that its sensitivity to its preferred stimulus is relatively **invariant** to the location in which it appears.

Perhaps even more provocative than the finding of an IT neuron with responses selective to a hand, was

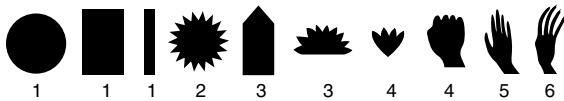


FIGURE 5.6 Examples of shapes that Gross, Rocha-Miranda, and Bender (1972) presented to a neuron in IT cortex, rank ordered by their ability to elicit activity in it, from none (1) or little (2 and 3) to maximum (6). Source: Gross, Rocha-Miranda, and Bender, 1972. Reproduced with permission of The American Physiological Society (APS).

Tip 5.4

In the visual system, and with other sensory cortex, the presentation of a preferred stimulus elicits a vigorous burst of action potentials. In the jargon of neurophysiology, this is often referred to as “driving” a response, or, even more informally, as “driving the cell.”

a subsequent report by this group of the results of recording from the superior bank of the superior temporal gyrus. Referring back to *Figure 5.5*, this corresponds to the regions labeled STPp (superior temporal polysensory – posterior), FST (**fundus** (or “floor”) of the superior temporal gyrus), and STPa (superior temporal polysensory – anterior). Gross and colleagues named this area the **superior temporal polysensory area (STP)** because many neurons in this area respond to stimuli from other sensory modalities in addition to visual. In one report, they found that nearly all neurons responded to visual stimuli, and of these, 21% responded to visual and auditory stimuli, 17% to visual and somesthetic stimuli, 17% to stimuli from all three modalities, and 41% exclusively to visual stimuli. The most provocative finding from this study was that a small number of neurons (7) was found that responded selectively to images of monkey and human faces. More specifically, these neurons seemed to respond to the category of *face*. For example, *Figure 5.7* illustrates a neuron that responded equally robustly to images of different monkeys’ faces, as well as to an image of a human face. Psychologists make the distinction between **type** (or category) of stimulus, such as the category of faces, versus a **token**, a specific exemplar drawn from within this category, such as the face of a particular person. Thus, the “face cells” discovered by Gross and colleagues seemed to be selective for stimuli of this particular type, but not for specific tokens within this category.

The two properties of the neurons of IT that we have emphasized here – a high degree of selectivity for complex stimuli, and large receptive fields – were in marked contrast to the known properties of neurons from V1. Neurons in V1 respond with high specificity to well-defined, low-level features, such as a boundary or line within a narrow range of orientation, or the end of a line. Additionally, within a module in V1 (i.e., within a pinwheel), there is an organization such that adjacent columns of neurons represent a smooth transition in the feature being represented. In temporal cortex, in contrast, the hand and face cells responded to the complete shape, but, for the most part, not to its component parts. Further, there was no evident topography to suggest that similar kinds of stimuli might be represented in adjacent areas of IT or STP. With regard to receptive fields, in V1, foveal receptive

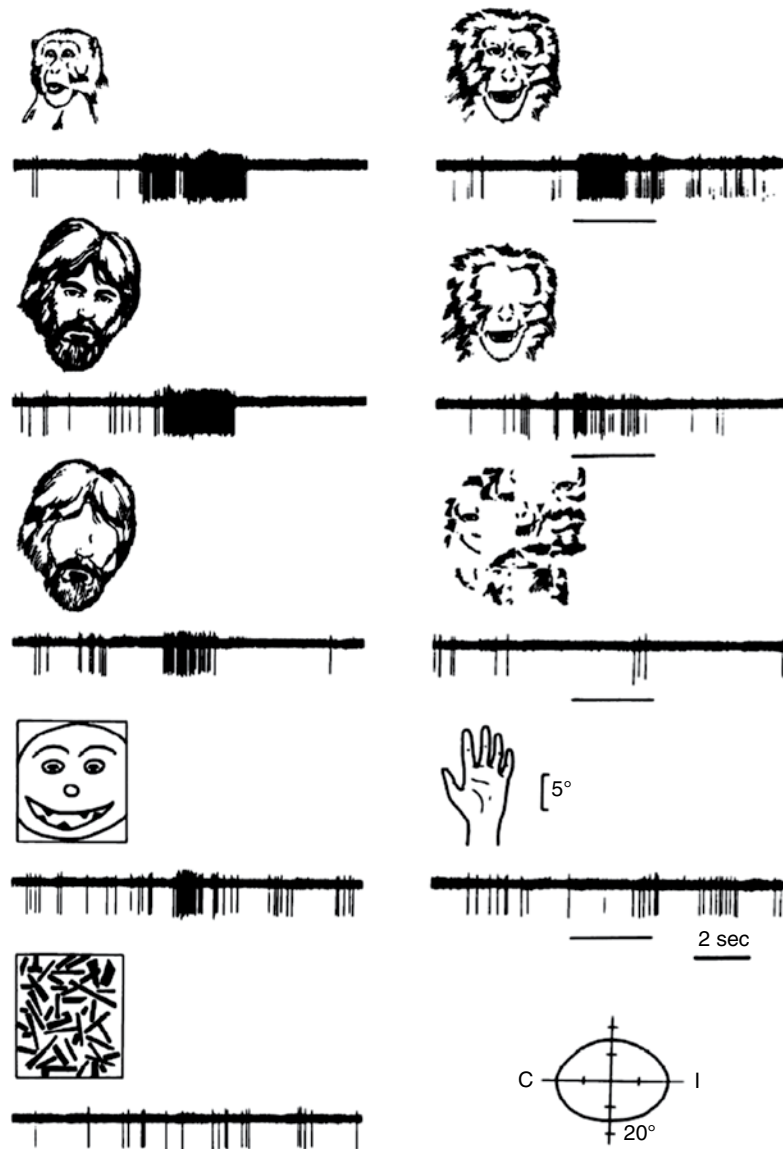


FIGURE 5.7 Responses of a face cell from area STP. Under each stimulus image is a trace from the electrode recording the neuron's activity, with each vertical bar corresponding to an action potential. Note that the neuron's response is comparably vigorous for the faces of two different monkeys and for a human face. The response is markedly weakened, though not abolished, when eyes are removed from the faces, as well as for an hand-drawn cartoon of a face. Changing the physical arrangement of the elements of an image, however, such as by scrambling the parts so that it no longer resembles a face, abolished the response. The lack of response to the hand is emblematic of the fact that this neuron did not respond to many other stimuli, ranging from bars, spots, and edges, to complex objects like hands and brushes. The lower right illustrates the neuron's receptive field (C=contralateral; I=ipsilateral). Source: Bruce, Desimone, and Gross, 1981. Reproduced with permission of The American Physiological Society (APS).

fields are quite small, on the order of $0.5^\circ \times 0.5^\circ$. The majority of neurons in temporal cortex, in contrast, had very large receptive fields. In STP, many of the receptive fields approached the size of the entire visual field. An additional feature that distinguished these temporal cortex neurons was that they could also be invariant to the size of their preferred image.

The invariance to location and to size of visually responsive temporal-lobe neurons had two implications. The first was that temporal cortex may contain the machinery to allow for the recognition of objects across a wide range of viewing conditions. This is a necessary property for vision in everyday life, because we rarely see an object under exactly the same viewing conditions on more than one occasion. For example, as I sit here typing this text, I recognize my mobile telephone lying on the desk to the left of my computer monitor. If I pick up the phone and bring it near my face, I will recognize it as the same unchanged, familiar object, this despite the fact that entirely different populations of cells in the retina, LGN, and V1 are processing low-level featural information about the object when it is in these two locations. Similarly, when my son walks into the room and turns on the light (I'm currently writing in my study at home, on an overcast day), I still recognize the telephone despite the fact that the colors that my retina detects differ considerably depending on the level of illumination in the room. The second implication was, in effect, a "problem" for the visual system that is created by these invariances. Note that when I look at my phone, pick it up, and set it down on the other side of my monitor, I not only recognize *what* it is but I also know *where* it is (*My phone is to the left of the monitor; My phone is near my face; etc.*). A mobile phone-tuned neuron with a large receptive field that spans the width of my computer monitor, however, won't carry information about where the phone is located. This follows from the fact that the response of this neuron is equivalent whether the phone is to the left of the monitor, to the right, or near my face. This sets up what has been referred to as the **binding problem**: *If ventral-stream neurons encode stimulus identity but not location, and if dorsal-stream neurons encode stimulus location, but not identity, how is this information brought together to create the unitary, "bound" percept that we experience of an object occupying a location in space?*

The binding problem is of central interest in many branches of cognitive neuroscience. Although we will not arrive at definitive solution to it in this chapter (nor, indeed, in this book), we will return to it periodically.

Broader implications of visual properties of temporal cortex neurons

In one sense, the studies by Gross and colleagues fit well with the findings from prior decades, that lesions of the temporal lobe produced profound deficits in the ability to recognize objects. They also brought to light many thorny questions about precisely how visual object recognition is accomplished by the brain. Recall, for example, that we noted that face cells seemed to respond to any token drawn from this category, rather than to a specific face. This raises two interesting puzzles. The more concrete of the two is: what supports our ability to recognize and discriminate among the tens of thousands of faces that we've encountered across our lives? This presents a puzzle because the neurons reported by Gross and colleagues didn't seem to have the specificity to support this ability. We will return to this in the next paragraph. The second question follows from the fact that these face cells seemed to signal the presence of an item from the category "face," rather than the presence of any specific face. Thus, in a sense, these neurons seemed to be conveying information about the idea of what it means to be a face – a face is the part of the body of many animals that includes the eyes, nose, and mouth; a face is associated with the individual identity of an animal more so than other parts of the body; a face conveys the mood or state of mind of the individual; and so on – rather than about the particular object being viewed. If the IT and STP are, indeed, parts of the visual object recognition system, how does one get from the low-level operations carried out in the retina, LGN, and V1 to cells in temporal cortex processing "the idea of a face"? Processing information about "the idea of a face," if that's what these neurons were doing, involves cognitive processes that are outside the domain of vision, strictly defined. Engaging knowledge about what a face is, what it can do, how a face is typically configured, and so on, falls under the rubric of semantic memory, a topic that we will address in detail in *Chapter 12*. That is, it entails accessing

knowledge about the world that was already present in the brain prior to the particular instance of seeing a face that occurred during that experimental session. These are deep questions that make explicit how the study of perception is central to building an understanding of cognition. For the time being, we'll leave it that the discovery of face cells made explicit the fact that a full understanding of visual object recognition will need to take into account the interface between the processing of newly encountered information in the world with our long-term store of knowledge about the world.

Let's return now to the fact that the face cells described by Gross and colleagues did not seem to show a specificity for any individual face. The failure to find any such neurons nonetheless required scientists to confront the implications of what it would mean if they *had* found such neurons. The expectation/assumption that there will be neurons that are selective for individual faces implies a visual system in which the processing of information at each stage becomes progressively more specific until, at the highest level, there is one neuron that represents each of the millions of objects in the world that we can recognize. Such a scheme for organizing the visual system, and the existence of highly specialized “gnostic” units (*Tip 5.5*) at the highest level, had been contemplated by the prominent Polish neuroscientist Jerzy Konorski (1903–1973). It had also been caricatured and disparaged by American neurophysiologist Jerome Lettvin (1920–2011) as the “grandmother cell theory.” The idea was that if there is a neuron specialized for recognizing every object with which one is familiar, among these must be a neuron that is specialized for recognizing one's grandmother.

Tip 5.5

The term “gnostic” has its roots in the Greek word for knowledge. The same etymological root was used to create the term visual agnosia (“absence of knowledge”), a syndrome characterized by the inability to grasp the meaning of an object, despite an intact ability to see it.

At first blush, the grandmother cell theory might have an intuitive appeal as the logical terminus of a hierarchical system of progressively increasing specialization. To be taken seriously, however, it would need to account for (at least) two major concerns. The first is vulnerability: if the system relies on highly specialized neurons at the apex of a processing hierarchy to represent each specific token that the organism can recognize, what would happen if one or more of these neurons were to be damaged? The second is the pragmatics of **phylogeny** and **ontogeny**: how, from the perspective of evolutionary or organismal development, could a system “know” a priori how many top-level gnostic units would need to be created and “set aside” such that the organism would have enough to represent each of the distinct objects that it would encounter throughout its life?

As we have seen, the research of Gross and colleagues in IT and STP provided some clues, and stimulated many ideas, about how visual object recognition might work. To make further progress however, it would be necessary to “close the gap” between the properties of neurons at these most advanced stages of the ventral visual pathway, and those of neurons at the earliest stages of cortical processing. Thus, it is to intermediate levels of the ventral stream that we turn next.

A hierarchy of stimulus representation

In the years between the early 1970s and today, many, many laboratories have performed experiments to “close the gap” between V1 and IT/STP. The picture that has emerged is generally consistent with the principle first introduced in *Chapter 3*: progressively higher levels of stimulus representation are constructed at progressively higher levels of the system by selective integration of more elemental information from earlier (or, in the terminology of a hierarchy, “lower”) stages of processing. Ungerleider and Bell (2011) liken this to the process of assembling words from syllables (*Figure 5.8*), which are, themselves, assembled from an even more elemental unit of written language, individual letters. In this analogy words correspond to high-level, complete object representations, syllables to intermediate-level conjunctions of

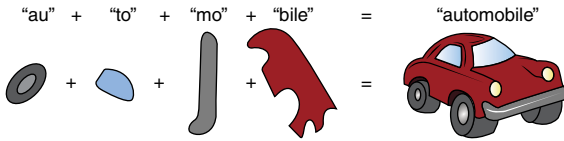


FIGURE 5.8 One analogy to object recognition is that just as words can be assembled from syllables, object representations may be assemblable from more primitive visual features. Source: Ungerleider and Bell, 2011. Reproduced with permission of Elsevier.

features, and letters to low-level features (such as those detected by V1). As detailed in *Chapter 3*, simple cells in V1 respond preferentially to specific luminance-defined orientations, and complex cells can signal the termination of lines. Neurons in area V2, located one synapse downstream, can respond to corners (which can be constructed from two differently oriented line segments). Neurons in V4, which receives projections from both V1 and V2, have a wide array of response profiles: some V4 neurons respond preferentially to curved contours (as illustrated in *Figure 5.9*); some to a specific combination of a degree of curvature at a particular orientation; some to outlines of basic geometric shapes; some to conjunctions of features, such as a line that is oriented at a particular angle *and* is of a particular color. At each of these stages, one sees that the stimulus information processed at that stage is more complicated than the information that was processed at the preceding stage(s) from which it receives feedforward inputs. Thus, one can say that neurons at higher stages in the ventral visual pathway become progressively more specialized. (For example, whereas a simple cell in V1 might respond to the “leading edge” of a bar tilted 45° to the right, regardless of color so long as the bar is darker than the background, a neuron in V4 might respond only to red bars of this orientation.) In most cases, one can see how the preferences of neurons at a particular stage of processing can be assembled, or, in some cases, inferred, from the convergence of inputs from an earlier stage. *Figure 5.9* offers an illustration of how this scheme of progressively more complex and specific stimulus representation might be engaged in the perception of a twisting tree branch. To return to the assembling-a-word analogy of visual perception (*Figure 5.8*), individual letters might

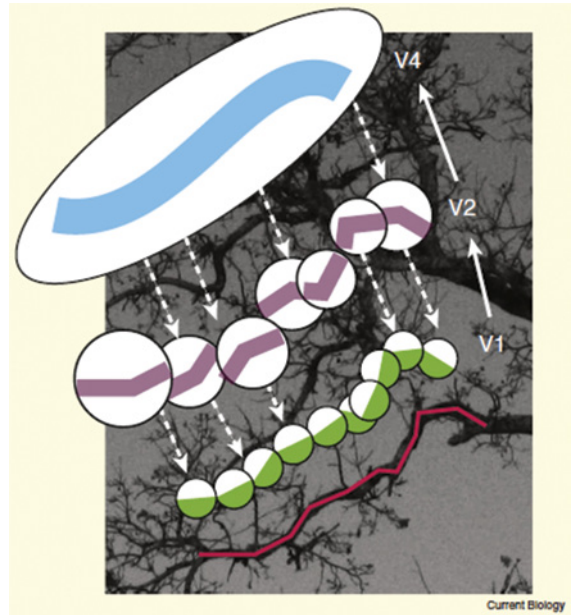


FIGURE 5.9 Hierarchical processing in the visual system. The perception of the branch of a tree (outlined with red line) proceeds from the fine-grained representation of the dense orientation structure by orientation-selective neurons in V1, to the representation of corners and straight sections of varying length in V2, to a representation of the overall contour of the branch in V4. Note that receptive fields increase in size at each level of processing. Solid arrows represent the feedforward direction of information processing, dashed arrows feedback. Source: Dakin, 2009. Reproduced with permission of Elsevier.

be represented in V1 and V2, syllables in V4, and the “final” percept, in this case an automobile, in IT.

Figure 5.9 also illustrates the fact that, in parallel to the trend of progressively narrowing selectivity of the visual features to which a neuron will respond, neurons at progressively downstream stages of processing in the ventral stream have progressively larger receptive fields. Thus, although a neuron in V4 will respond to many fewer stimuli than will a neuron in V1, the V4 neuron will respond to its preferred stimuli when they appear in a much larger range of the visual field. As we saw in *Chapter 3* and *Chapter 4*, this is a necessary byproduct of integrating over several smaller receptive fields from a lower level. It is also important

to note one of the efficiencies of a feature-based hierarchical object recognition system, which is that the lower levels of the system can contribute to the perception of many, many objects. Returning again to the assemble-a-word analogy (*Figure 5.8*), if the word being presented were *apple*, instead of *automobile*, the “a” and “e” feature detectors (from our fanciful V1) could also contribute to the perception of this very different word. For *apple*, the *a* would feed forward to a neuron in V4 representing the syllable *app*, and the *e* detector to a neuron in V4 representing the syllable *le*. Neurons at the level of our fanciful area V4 could also contribute to the perception of multiple words, the *mo* detector, for example, also contributing to the perception of *motion*, or *dynamo*.

Paradoxically, however, what makes for an efficient system at lower levels might lead to a hopelessly inefficient, and perhaps impossible, situation at the highest level. That is, the logical end of a hierarchical chain of progressively more specialized feature detectors is a highest level of representation at which individual neurons are each dedicated to one of the seemingly infinite number of objects that we can recognize – one for an *automobile*, one for an *apple*, one for *motion*, one for a *dynamo*, and so on, ad infinitum. The existence of such a “grandmother level” at the top of the visual perception hierarchy poses at least two problems. The first is the vulnerability of such a system. As hinted at before, damage to a large swath of the “grandmother level” would render one unable to recognize many familiar objects. The second problem is one of storage. In the infant, how could one know, a priori, how many neurons to “set aside” for every possible object that the organism might encounter over the course of its life? And although we know that the cerebral cortex contains on the order of 30 billion neurons, would this be enough, when one takes into account all the other functions that it must support, to represent all the objects that we will perceive in our lives on a one-neuron-to-one-object basis? (Note that, hearkening back to *Thought Box 1.2*, the existence of a level of “grandmother cells” can be considered an extreme version of localization of function.)

One solution to the problems voiced here is to devise a scheme for object representation that does not rely on a “grandmother level,” and that’s what we consider next.

Combination coding, an alternative to the “grandmother cell”

A breakthrough in thinking about how object recognition might be accomplished at the highest levels came from Japanese neurophysiologist Keiji Tanaka and his colleagues at the Institute of Physical and Chemical Research (RIKEN). The key to their breakthrough came from their “reduction method” of studying stimulus selectivity in IT neurons. They made their recordings in anterior IT cortex, the region labeled TE in *Figure 5.4*, and CITv and AITv (central IT, ventral portion; anterior IT, ventral portion) in *Figure 5.5*. Upon isolating a single neuron with a microelectrode, they would first present the monkey with dozens of three-dimensional objects, drawn from many categories, so as to find the complex stimulus that was most effective at driving the cell. Having done so, they next created a 2-D facsimile of the 3-D object via the superposition many simple features. *Figure 5.10* illustrates this process for an IT neuron that was discovered to be most responsive to a top-down view of a tiger’s head. (Presumably, in the first step of the process, the investigators presented a toy tiger and the neuron responded vigorously. Upon holding the tiger at different distances, and rotating it to present different views, they determined that the strongest responses were elicited by displaying the head from this top-down view.) The key to the 2-D facsimile image is that it was assembled to be “decomposable” into progressively simpler configurations. For the tiger head, for example, the 2-D facsimile image consisted of an oval (a stand-in for the shape of the head), a superimposed array of a texture of black and white bars (“stripes”), and two black rectangles (stand-ins for the ears). Critically, the investigators confirmed that the neuron’s response to this 2-D facsimile was comparable to its response to the tiger head itself. The “reduction” process, then, proceeded via the progressive removal, or in some cases simplification, of individual elements of the facsimile image, until reaching the point at which the neuron’s response was no longer as robust as it had been to the original 3-D object.

For the IT neuron featured in *Figure 5.10*, this procedure revealed that this neuron’s true selectivity was not for a top-down view of a tiger’s head (which would be a highly selective feature, indeed!) but, rather, for a simpler configuration of a white square

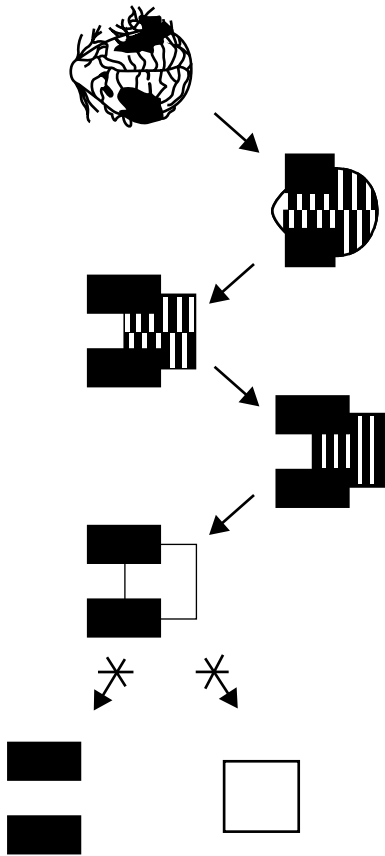


FIGURE 5.10 Tanaka's reduction method for determining the tuning properties of an IT neuron. After the neuron was determined to be maximally responsive to the top-down view of a tiger's head (top image), the investigators confirmed that it responded equally vigorously to a 2-D facsimile image comprising an oval, a pattern of black and white bars, and two black rectangles. The facsimile image was then "reduced" in a step-by-step manner until the investigators reached the simplest configuration of elements beyond which they could not further decompose the image without eliminating the neuron's response. Source: Tanaka, 1993. Reproduced with permission of AAAS.

with two superimposed black rectangles. As the investigators heroically repeated this process with hundreds of real-world 3-D objects, they confirmed, again and again, that IT neurons that were initially found to respond vigorously to one of these objects were, in fact, tuned to a simpler, reduced set of generic features (Figure 5.11 shows an example of many of these).

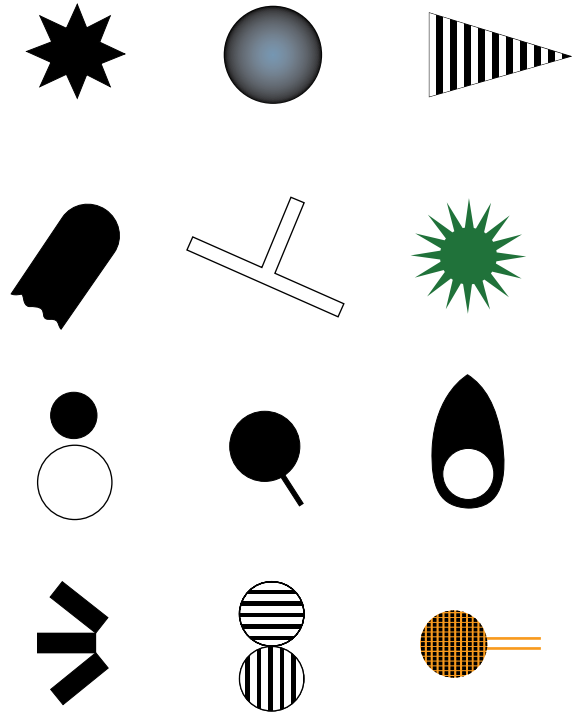


FIGURE 5.11 Twelve examples of the critical "feature" (i.e., the image identified via the reduction method), each found to be effective for driving a neuron in anterior IT. Source: Tanaka, 1993. Reproduced with permission of AAAS.

Thus, the visual system may not accomplish the recognition of a top-down view of a tiger's head via a hierarchy of progressively more specialized neurons that terminates with a "top-down-view-of-a-tiger-head cell" in a putative "grandmother cell" level at the apex of this hierarchy. Instead, the tiger's head may be recognized by the simultaneous (i.e., the "combined") activation of several tens (if not hundreds or more) of neurons that each represent a different set of features that uniquely make up this stimulus: one group of cells representing the configuration of two black rectangles superimposed on one white square, as illustrated in Figure 5.10, another group representing the black and white stripes, another group representing the moist, glistening skin of the nose, another group representing the whiskers, and so on. Note that, in contrast to the extreme localizationism implied by a hierarchy ending in a "grandmother level," the combination coding proposed by Tanaka and colleagues

has two hallmarks of a distributed system. First, there is no one place where a tiger head is represented. Instead, its representation entails the simultaneous activation of many elements that are distributed throughout IT cortex. Second, any one of these elements can participate in the representation of many different objects. For example, although the recognition of a top-down view of the head of an otter or a warthog is sure to activate many elements not activated by the tiger, it may nonetheless activate the same IT neurons that can represent whiskers.

There are two additional points raised by the discoveries of Tanaka and colleagues that we will consider here. The first, illustrated in *Figure 5.12*, is that there does appear to be a columnar organization to IT, albeit according to a more complex scheme than that found in V1. The second is that faces were the one class of stimuli that Tanaka and colleagues found that were generally not amenable to decomposition by the reduction method. That is, when an IT neuron was found that responded vigorously to a face stimulus, it was generally not possible to decompose the image into a simpler collection of features that would still drive the cell equally well. (The one exception was a single neuron described in one study for which a face could be simplified to two eyes and a nose.) This finding was generally consistent with what Gross and colleagues had reported roughly 20 years earlier (see *Figure 5.7*).

Research Spotlight 5.2 highlights a particularly clever and technically challenging experiment from Wang, Tanaka, and Tanifuji (1996), combining extracellular electrophysiology and optical imaging, that provided considerable insight about the columnar organization of IT. It also re-raises the idea, briefly considered earlier in this chapter, that one function of anterior IT may be to represent faces as a type of object, as opposed to the discriminating and recognizing exemplar faces. Thus, again, we see evidence for the blurring of the lines between perception and memory in anterior IT. Finally, consideration of this study also invokes the distinction, pursued in *Thought Box 5.2*, between two modes of object recognition: feature-based vs. image-based. (*Web Link 5.1* illustrates how these principles are manifest in neuroimaging studies of human object recognition.)

A critical role for feedback in the ventral visual processing stream

Visual illusions provide evidence for top-down influences

A final property of the ventral stream that we will consider in this chapter is its bi-directional interactivity, an extension of the property introduced in *Chapter 3*. An easy way to illustrate this is with visual illusions, which often “work” because our visual system often “makes inferences” about a scene, such as

FIGURE 5.12 Columnar organization of IT.

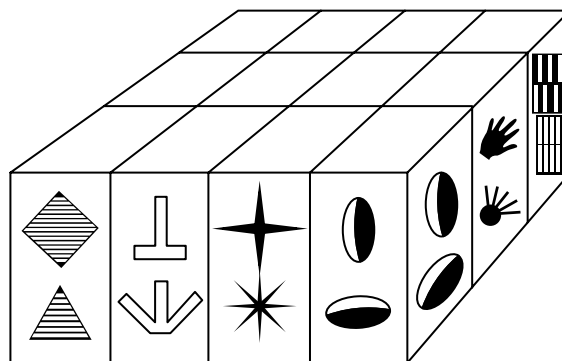


FIGURE 5.12.A Electrophysiological and optical imaging experiments from Tanaka’s group have suggested a columnar organization in IT, as cartooned here. One characteristic that they discovered is that different neurons within a column often represented similar, but not identical, features. Source: Tanaka, 1993. Reproduced with permission of AAAS.

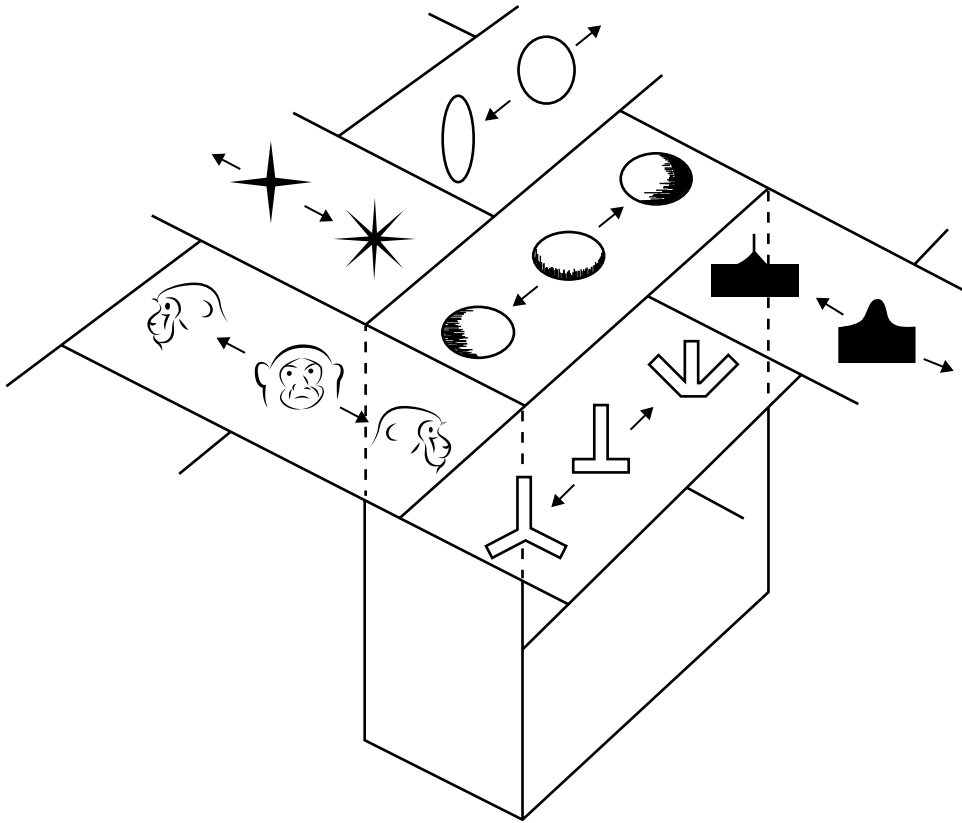


FIGURE 5.12.B As would be necessitated by a high-dimensional feature space (i.e., IT must represent color, degree of oval-ness vs. circle-ness, width of bars in alternating dark-light grating, number of arms in a “star” shape, etc.), whereas some adjacent columns in IT represent related features, there are also often discontinuities. Nonetheless, as revealed by the study highlighted in *Research Spotlight 5.2*, there is evidence of continuous mapping of some stimulus properties, such as of a face rotating from left-profile view, through “full frontal” view, to right profile view. Source: Tanaka, 1997. Reproduced with permission of Elsevier.

RESEARCH SPOTLIGHT

5.2 The continuous representation of rotated views of a face in IT (Wang, Tanaka, and Tanifuji, 1996)

Figure 5.13 illustrates how, in order to explore the columnar organization of IT, Wang, Tanaka, and Tanifuji (1996) used a procedure of first performing extracellular recordings, to determine the tuning properties of one or more neurons in a region, then presenting the same preferred stimulus (plus non-preferred control stimuli) while acquiring optical images from the same region of cortex. The top row of *Figure 5.13* shows

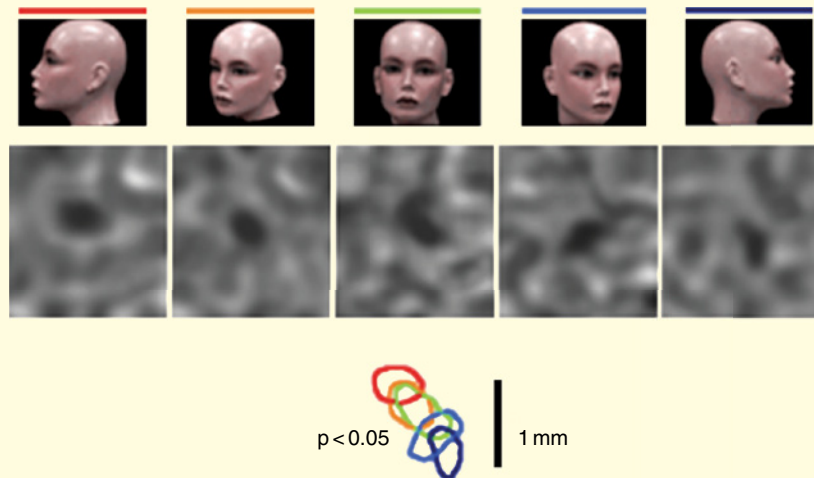
Research Spotlight 5.2 *Continued*

FIGURE 5.13 Exploring the columnar organization of IT. Source: Figure courtesy of Keiji Tanaka, modified from Wang, Tanaka, and Tanifuji, 1996. Reproduced with permission of AAAS.

each of five views of a doll's head presented to a monkey. The stripe of color above each identifies the corresponding outline of the associated region of elevated activity that is illustrated in the lower portion of the figure. The electrode penetration that initiated this session found five neurons that responded selectively to faces, three to a full-face view, and two to profile views. The middle row shows the optical signal, each of the same patch of cortex, associated with each of the five views. These images, acquired without dye, are of reflected light at 605 nm wavelength. The intensity of this reflected light decreases with increased neuronal activity (due to local changes in hemoglobin oxygenation). Thus, dark regions in these images are regions of greater neuronal activity (see *Methodology Box 3.1*). The bottom row shows outlines of the central dark spot from each of the five views, thresholded at $p < .05$, color-coded, and superimposed to show the dark spot's relative position for each view. Data from a second monkey (not shown here) revealed very similar patterns of activity when this animal viewed either the doll's head in these five orientations, or the head of one of the experimenters in the same five orientations. One implication of these findings is that, as cartooned in *Figure 5.12.B*, local regions of IT may represent the same object as viewed from different perspectives. Possible implications of this are considered in *Thought Box 5.2*.

THOUGHT BOX

5.2 Object-based (viewpoint-independent) vs. image-based (viewpoint-dependent) representation in IT

Implied in the *assemble-a-word* analogy to object recognition (Figure 5.8) is the assumption that object representations are viewpoint-independent. That is, it need not matter whether one is viewing, say, an automobile, from the front, side, or back; so long as one can identify enough constituent parts (e.g., one or more tires, a bumper, the red, curved metal of a part of the body), one can identify it as an automobile. The visual system would seem to be well constructed to perform such viewpoint-independent – or object-based – recognition. For example, whether viewing the car from the side, and therefore the curved chrome shape of the bumper appears in the left visual field, or from the front, with bumper occupying central vision just below the fovea – both regions of the visual field are covered by V1 feature detectors that will detect and pass along information about the contour and color of the bumper to higher levels of the system that will, progressively, complete the process of recognition. Although we've already discussed how the work of Tanaka and colleagues helped to dispel the idea that there may exist a “grandmother level” of representation, such a model nevertheless implies that there exists an abstract representation of “automobile” that is, well, independent of any specific angle from which it can be viewed.

There is, however, evidence consistent with alternative, *viewpoint-dependent* encoding of objects in IT. Behaviorally, it has been shown that people are faster, and more accurate, at recognizing objects from previously seen than from previously unseen orientations. This has been done, for example, with “bent paperclip” shapes that subjects would never have seen prior to participating in the experiment. Using these shapes, neurophysiologist Nikos Logothetis and colleagues have recorded from neurons in IT cortex while monkeys viewed the same type of stimuli. An intriguing finding from these studies is that some IT neurons showed their strongest response to a particular preferred viewing angle of a novel object, and these responses dropped off monotonically as the object rotated further from this preferred view. Considering these findings makes one think that the columnar organization of tuning for faces-in-profile (Figure 5.13) might also reflect image-based representation.

It is obvious from the Logothetis studies (reviewed in Logothetis and Sheinberg, 1996), because they were performed with stimuli that their animals had never seen prior to the beginning of the experiment, that the neurons' viewpoint preferences with these stimuli were acquired through experience. Thus, viewpoint-dependence for certain stimuli may be a property that can be acquired through experience. Indeed, this may be the basis for some forms of expertise. Having made these points, however, one must also consider the fact that, early in development, the establishment of viewpoint-dependent representations *must* precede that of viewpoint-independent representations, for the simple reason that our first encounter with any object is necessarily from a particular point of view. Indeed, it would seem that a viewpoint-independent representation can't be constructed until a minimum number of distinct views has been acquired. The extent to which viewpoint-dependent and -independent representations co-exist, and interact, remains an important question for the cognitive neuroscience of object recognition.

assuming the presence of an element that is not actually there. In *Figure 5.14*, for example, most people see a white triangle in the foreground of image 5.14.A (equilateral, with one side parallel to the bottom of

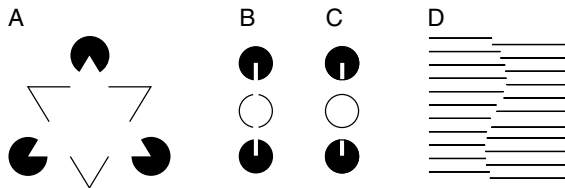


FIGURE 5.14 Sensitivity to illusory contours in V2. **A**, **B**, and **D** illustrate illusions in which many report seeing a white triangle (A), a vertical white line (B), or a curving boundary between two surfaces (D). **C** illustrates how just a small modification to the image can remove the illusion. Source: Von der Heydt, Peterhans, and Baumgartner, 1984. Reproduced with permission of AAAS.

the page) when, in fact, there is no triangle. In effect, the visual system seems to “draw” white lines where none actually exist. (Mine actually “paints in” an entire 2-D surface.) Insight into why the visual system does this can be gained from comparing the images in 5.14.B and 5.14.C. In 5.14.B, “local” details from each of the three circles are consistent with there being a white bar lying across all three, and so our visual system “draws one in”; in 5.14.C, in contrast, the same local regions do not contain this evidence, and so no bar is perceived.

Is there any neural reality to the sense that one gets that, in *Figure 5.14.A* and *B*, white lines are “being drawn” to create the percept of an object that isn’t really there? A highly influential study performed in Zurich by the physiologist and theoretician Rüdiger von der Heydt and colleagues suggests that there may be. As illustrated in *Figure 5.15*, the researchers, recording from neurons in area V2, first found the feature to

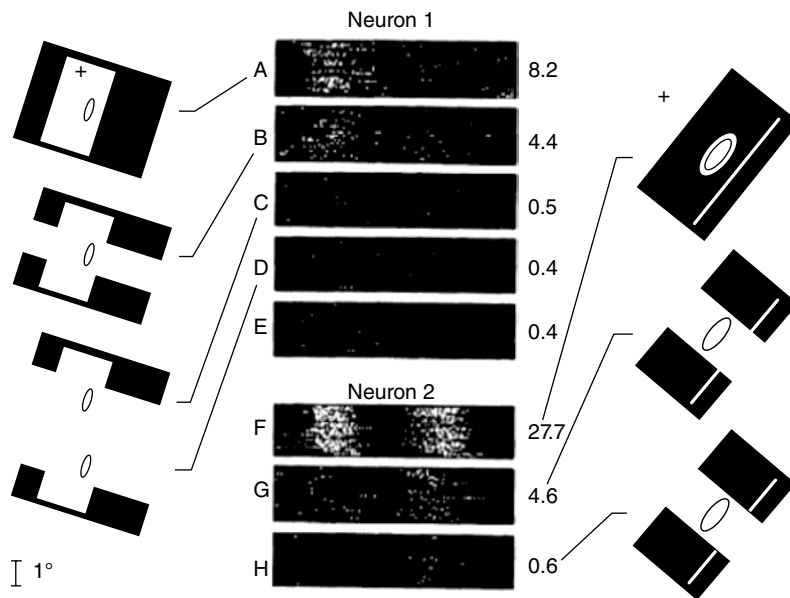


FIGURE 5.15 The responses of two V2 neurons reported by von der Heydt, Peterhans, and Baumgartner (1984; center column), with the stimuli producing these responses shown to the left (for neuron 1) and to the right (neuron 2). For both neurons, the receptive fields are represented by ellipses. Neuron 1 responds most strongly to the right-side edge of an oriented bar (raster plot A), but also to the analogous illusory edge (raster plot B), a response that is abolished when either the top or bottom half of the stimulus is removed (raster plots C and D). Raster plot E shows spontaneous activity. Neuron 2 responds most strongly to a narrow oriented bar (raster plot F), less strongly to the illusory bar (raster plot G), and effectively not at all to a minor modification that abolishes the illusory percept (raster plot H). Source: Von der Heydt, Peterhans, and Baumgartner, 1984. Reproduced with permission of AAAS.

which the neuron was tuned (the edge of a tilted, fat white bar in 5.15.A, and a thin white bar in 5.15.F). Next, they modified the stimuli to construct an “illusory contour” where the real contour had been (5.15.B and G) and swept the modified stimulus through the visual field of the neuron, being careful that only the illusory part passed through the receptive field. Remarkably, for the two neurons illustrated in panels 5.15.A–E and 5.15.F–G, the V2 neurons responded in a weaker, but nonetheless selective, way. The specificity of the illusory effect was shown when, in a manipulation analogous to what we saw in Figure 5.14 with display C vs. display B, the neurons did not respond to a third condition in which local changes were made that abolished the percept of the illusion.

How are we to understand this result? First, as would also be the case with the white triangle and white bars from Figure 5.14, the “objects” created by the illusions are considerably larger than the V2 receptive fields from which von der Heydt and colleagues recorded. That, plus what we’ve already established about the ventral-stream hierarchy, means that the representation of the “fat bar” in Figure 5.15.B and of the “thin bar” in 5.15.G were almost certainly formed in V4, if not in posterior IT. It must be the case, therefore, that the “inference” that there was an object in

the visual field was fed back from V4 (or IT) to V2, and the responses of the V2 neuron were a consequence of this top-down signal. Von der Heydt, Peterhans, and Baumgartner (1984) provided even more compelling evidence for this account with data from a third neuron recorded from V2 (Tip 5.6). As illustrated and described in Figure 5.16, this neuron responded to the illusory vertical line created by the interface of two offset gratings (analogous to the illusory “curve” in Figure 5.14.D). To simulate what the experimenters did, begin by holding a blank piece of paper such that you are blocking all but the bottom

Tip 5.6

It’s very rare that one would be able to publish a paper reporting the findings from just three neurons in a peer-reviewed journal, not to mention in *Science*! For von der Heydt, Peterhans, and Baumgartner (1984), however, it is evidently the case that the elegance of the experimental design, the painstaking nature of the data collection process, and the blindingly clear implications of the results were enough to exceed the journal’s notoriously high threshold.

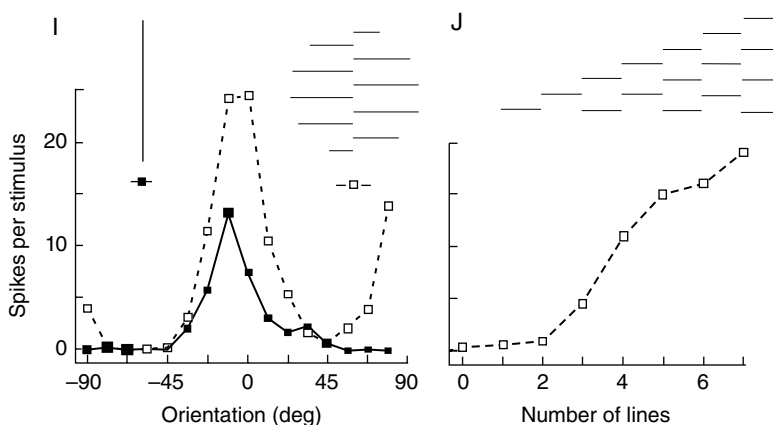


FIGURE 5.16 Data from neuron 3 of von der Heydt, Peterhans, and Baumgartner (1984). Panel I shows that the tuning of this V2 neuron is to a (real) black bar oriented nearly vertically (solid black lines on plot). The dashed lines show the cell’s tuning to the illusory line created by the stimulus to the right is nearly identical (although the response gain is higher for the illusory than the “real” stimulus). Panel J shows the sensitivity of this neuron to the strength of the illusory percept, its responses increasing as a function of the number of lines added to the gratings in the stimulus. Source: Von der Heydt, Peterhans, and Baumgartner, 1984. Reproduced with permission of AAAS.

line of the grating on the left side of the “circle” formed by these two gratings. From this starting point, there can be no vertical line because none of the opposing bars from the grating on the right side of the image are visible. Next, slowly slide your blank piece of paper upwards, so that you progressively reveal each bar in the image. As you get beyond the third bar, the percept of the illusory line will start to feel stronger and stronger. What panel *J* in *Figure 5.16* illustrates is that, in parallel to this progressive strengthening of the illusory percept, the V2 neuron emitted a progressively stronger firing rate.

Top-down feedback can disambiguate the visual scene

Although the multiplicity of small receptive fields in the retina, LGN, and V1 that “tile” the visual field offer high acuity, they also create challenges for the visual system, one of which is known as the **aperture problem**. At a generic level, what this refers to is the fact that each neuron with a small receptive field is, in effect, viewing the visual scene through a very small aperture. (Imagine looking through the peephole in a door if the fish-eye lens has been removed from it.) *Figure 5.17* illustrates this problem for motion perception: when an irregularly shaped object is moving in one direction, the direction of motion of small portions of the edge of the object, when seen through a small aperture, can be ambiguous. To illustrate, the figure shows one receptive field on the neck of the deer at which at least three direction-sensitive feature detectors (each tuned for one of the directions indicated by the red and green arrows) might signal that its preferred direction of motion is occurring within this receptive field. Although this would result in a noisy, ambiguous message being sent “up the chain” to MT, we know from *Chapter 3* that there is circuitry in place whereby MT can control activity in V1 and the LGN, thereby reducing this feedforward cacophony and emphasizing the motion signals that it would “prefer” to receive (see *Figure 3.6*). Indeed, one of the points made in that earlier discussion was that, in the same way that MT can control V1 and the LGN, there must be a higher level of representation feeding back onto MT. We are now in a position to consider what that higher level of representation might be.

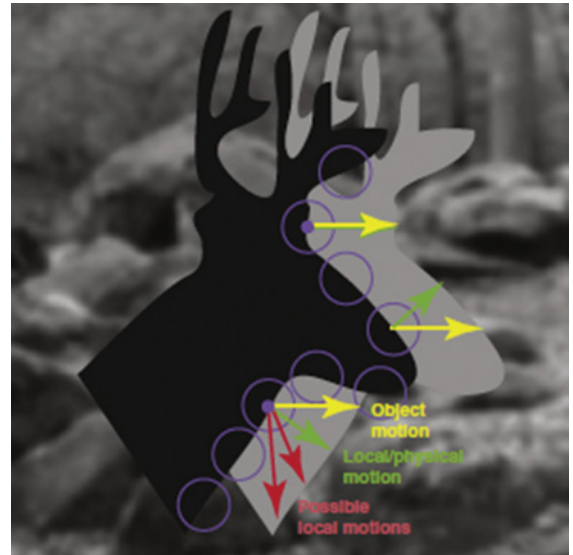


FIGURE 5.17 An example of how high-level visual object recognition is needed to influence earlier stages of visual processing. The black silhouette of the deer head represents its current location, and the gray silhouette a location it will occupy in the very near future. The purple circles represent V1 receptive fields, with yellow arrows indicating the true direction of motion of the object (i.e., of the deer), green arrows the physical motion within two receptive field “apertures,” and the red arrows two other possible interpretations of the direction of motion within one of these receptive fields. Source: Dakin, 2009. Reproduced with permission of Elsevier.

Recall from *Figure 5.5.B* that there are lateral interconnections between MT and V4. In our example of the running deer, then, networks in IT that have assembled the representation of the deer can feed this information back onto V4 (as well as, perhaps, laterally to area MST), V4 can then share it with MT, and MT is now equipped to “know” how to control the bottom-up flow of motion signals coming from V1.

TAKING STOCK

Although there is so, so much more to know about visual object recognition and the functioning of the ventral visual stream, my editor is already getting

uneasy about the length of this chapter. Therefore, let's just take a moment to review what we've covered and, in particular, how it relates to what's yet to come. The anatomical connectivity of the visual system, beginning at the retina if one considers the M and P classes of RGCs, creates a substrate for two broad classes of visual information processing: determining *what* an object is, and determining *where* it is. (And we've already noted that a question for *Chapter 7* is whether *where* is perhaps better construed as *how*.) Lesion studies have confirmed, in the form of a double dissociation of function, that the temporal lobe, particularly IT cortex, contributes importantly to object recognition, and the PPC contributes to spatial processing. This chapter's focus on the ventral stream has emphasized the hierarchical nature of object recognition. After flirting with a grandmother layer of object representation, a distributed, combination-coding scheme was seen to offer a more plausible account of stimulus processing in IT cortex. Particularly intriguing is the neural processing of faces. For one thing, faces don't display "neural decomposition" as readily as do other complex objects. Additionally, IT responses to faces suggest two interesting properties of this system. First, IT may support both object-based and

image-based representations. Second, some neurons in IT may represent "the idea of a face," rather than any one exemplar face.

This latter possibility presages a return visit to ventral temporal cortex. After the next three chapters conclude this section's emphasis on *perception*, *attention*, and *action*, *Chapter 9* will address the question (with apologies to *Chapter 3*) *Where does perception end, and where does memory begin?* As with the earlier question (which was posed about sensation vs. perception), we'll see that the boundary is a blurry one, one having more basis in arbitrary labeling than in neurobiological or psychological principle. For example, this chapter will consider agnosias (see *Tip 5.5*), syndromes of "perception stripped of its meaning" that result from damage to the ventral temporal lobe. It will consider the phenomenon of "category specificity" in ventral temporal cortex (e.g., the *fusiform face area*, the *parahippocampal place area*, the *extrastriate body area*, and the *visual word form area*): how did these areas come into being (nature vs. nurture); what does their existence tell us (or not) about the organization of knowledge in the brain; and, more concretely, what are the limits of what we can infer from activity in one of these areas? (To whet your appetite, check out *Web Clip 5.1*.)

END-OF-CHAPTER QUESTIONS

1. In what sense does function follow from structure in the visual system?
2. Design an experiment that, if your hypothesis is correct, might produce evidence for a double dissociation of function. In what way would your hoped-for results support stronger inference than a single dissociation?
3. Define the functions of the *what* and *where* pathways. Summarize at least three pieces of evidence for this organizing principle of the visual system.
4. Summarize the type of image processing that occurs at each level of the ventral-stream hierarchy.
5. Define the concept of the "grandmother cell." For what reasons is it logically appealing? For what reasons, practical or theoretical, is it problematic?
6. What is a plausible alternative to a "grandmother level" of object recognition? Summarize the evidence for it.
7. Are faces processed differently than other objects by the visual system?
8. Define object-based vs. image-based models of object recognition. Name an advantage, and a limitation, of each.
9. How does feedback from temporal cortex play an important role in visual-scene processing?

REFERENCES

- Born, Richard T. 2001. "Visual Processing: Parallel-er and Parallel-er." *Current Biology* 11 (14): R566–R568. doi: 10.1016/S0960-9822(01)00345-1.
- Bruce, Charles, Robert Desimone, and Charles G. Gross. 1981. "Visual Properties of Neurons in a Polysensory Area in Superior Temporal Sulcus of the Macaque." *Journal of Neurophysiology* 46 (2): 369–384.
- Curtis, C. E., and D'Esposito, M. 2004. "The Effects of Prefrontal Lesions on Working Memory Performance and Theory." *Cognitive, Affective, and Behavioral Neuroscience* 4(4): 528–539.
- Dakin, Steven C. 2009. "Vision: Thinking Globally, Acting Locally." *Current Biology* 19 (18): R851–R854. doi: 10.1016/j.cub.2009.08.021.
- Felleman, Daniel J., and David C. Van Essen. 1991. "Distributed Hierarchical Processing in Primate Visual Cortex." *Cerebral Cortex* 1 (1): 1–47. doi: 10.1093/cercor/1.1.1.
- Gross, Charles G., C. E. Rocha-Miranda, and D. B. Bender. 1972. "Visual Properties of Neurons in Inferotemporal Cortex of the Macaque." *Journal of Neurophysiology* 35 (1): 96–111.
- Logothetis, Nikos K., and David L. Sheinberg. 1996. "Visual Object Recognition." *Annual Review of Neuroscience* 19: 577–621. doi: 10.1146/annurev.ne.19.030196.003045.
- Mishkin, Mortimer, Leslie G. Ungerleider, and Kathleen A. Macko. 1983. "Object Vision and Spatial Vision: Two Cortical Pathways." *Trends in Neurosciences* 6: 414–417. doi: 10.1016/0166-2236(83)90190-X.
- Pohl, Walter. 1973. "Dissociation of Spatial Discrimination Deficits Following Frontal and Parietal Lesions in Monkeys." *Journal of Comparative and Physiological Psychology* 82 (2): 227–239. doi: 10.1037/h0033922.
- Tanaka, Keiji. 1993. "Neuronal Mechanisms of Object Recognition." *Science* 262 (5134): 685–688. doi: 10.1126/science.8235589.
- Von der Heydt, Rüdiger, Ernst Peterhans, and G. Baumgartner. 1984. "Illusory Contours and Cortical Neuron Responses." *Science* 224 (4654): 1260–1262.
- Wang, Gang, Keiji Tanaka, and Manabu Tanifuji. 1996. "Optical Imaging of Functional Organization in the Monkey Inferotemporal Cortex." *Science* 272 (5268): 1665–1668.

OTHER SOURCES USED

- Blake, Randolph, and Robert Sekuler. 1985. *Perception*. New York: Alfred A. Knopf.
- Gross, Charles G., D. B. Bender, and C. E. Rocha-Miranda. 1969. "Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey." *Science* 166 (3910): 1303–1306. doi: 10.1126/science.166.3910.1303.
- Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell, eds. 2000. *Principles of Neural Science*. New York: McGraw-Hill.
- Schiller, Peter, H. 1993. "The Effects of V4 and Middle Temporal (MT) Area Lesions on Visual Performance in the Rhesus Monkey." *Visual Neuroscience* 10 (4): 717–746. doi: 10.1017/S0952523800005423.
- Schiller, Peter H., and Nikos K. Logothetis. 1990. "The Color-Opponent and Broad-Band Channels of the Primate Visual System." *Trends in Neurosciences* 13 (10): 392–398. doi: 10.1016/0166-2236(90)90117-S.
- Semmes, Josephine, Sidney Weinstein, L. Ghent, John S. Meyer, and Hans-Lukas Teuber. 1963. "Correlates of Impaired Orientation in Personal and Extrapersonal Space." *Brain* 86 (4): 747–772.
- Tanaka, Keiji. 1996. "Inferotemporal Cortex and Object Vision." *Annual review of Neuroscience* 19: 109–139. doi: 10.1146/annurev.ne.19.030196.000545.
- Tanaka, Keiji. 1997. "Mechanisms of Visual Object Recognition: Monkey and Human Studies." *Current Opinion in Neurobiology* 7 (4): 523–529. doi: 10.1016/S0959-4388(97)80032-3.
- Ungerleider, Leslie G., and Mortimer Mishkin. 1982. "Two Cortical Visual Systems." In *Analysis of Visual Behavior*, edited by David J. Ingle, Melvin A. Goodale, and Richard J. W. Mansfield, 549–586. Cambridge, MA: MIT Press.
- Ungerleider, Leslie G., and Anderw H. Bell. 2011. "Uncovering the Visual 'Alphabet': Advances in Our Understanding of Object Perception." *Vision Research* 51 (7): 782–799. doi: 10.1016/j.visres.2010.10.002.
- Wolfe, Jeremy M., Keith R. Kleunder, Dennis M. Levi, Linda M. Bartoshuk, Rachel S. Herz, Roberta Klatzky, ... Daniel M. Merfeld. 2012. *Sensation and Perception*, 3rd ed. Sunderland, MA: Sinauer Associates.

FURTHER READING

Logothetis, Nikos K., and David L. Sheinberg. 1996. "Visual Object Recognition." *Annual Review of Neuroscience* 19: 577–621. doi: 10.1146/annurev.ne.19.030196.003045.

A very extensive review emphasizing behavioral, neuropsychological, and electrophysiological evidence for viewpoint-dependent mechanisms of object recognition, as well as for different populations of IT neurons specialized to perform different functions, such as categorization, recognizing living things, and supporting the action of object grasping.

Mishkin, Mortimer, Leslie G. Ungerleider, and Kathleen A. Macko. 1983. "Object Vision and Spatial Vision: Two Cortical Pathways." *Trends in Neurosciences* 6: 414–417. doi: 10.1016/0166-2236(83)90190-X.

A concise summary of the two-parallel-pathways model of the visual system.

Schiller, Peter H., and Nikos K. Logothetis. 1990. "The Color-Opponent and Broad-Band Channels of the Primate Visual System." *Trends in Neurosciences* 13 (10): 392–398. doi: 10.1016/0166-2236(90)90117-S.

A thorough summary, and synthesis, of the relative contributions of the M and P channels to visual perception.

Tanaka, Keiji. 1996. "Inferotemporal Cortex and Object Vision." *Annual Review of Neuroscience* 19:109–139. doi: 10.1146/annurev.ne.19.030196.000545.

An overview of this investigator's pioneering work on the neural principles underlying object recognition in IT cortex.

Ungerleider, Leslie G., and Anderw H. Bell. 2011. "Uncovering the Visual "Alphabet": Advances in Our Understanding of Object Perception." *Vision Research* 51 (7): 782–799. doi: 10.1016/j.visres.2010.10.002.

A remarkably comprehensive overview of the cognitive neuroscience of object recognition, ranging from nineteenth-century studies of brain and behavior (some of which we considered in Chapter 1) to an insightful critique of contemporary methods for multivariate pattern analysis of fMRI data (which we will introduce in Chapter 8).

CHAPTER 6

SPATIAL

COGNITION

AND ATTENTION

KEY THEMES

- Neuropsychology can be an important tool for hypothesis generation, as evidenced by the models of attention derived from the study of unilateral neglect.
- The neglect syndrome provides evidence that attention: (a) can operate within an egocentric frame of reference; (b) can be location-based or object-based; (c) may operate differently in peripersonal vs. extrapersonal space; and (d) may act to resolve competition among neural representations.
- The cortical regions whose damage produces unilateral neglect have been studied in neurologically healthy individuals with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), and show elevated activity during tasks requiring spatial- vs. object-based visual discriminations, as well as retinotopic organization.
- Neurons of posterior parietal cortex (PPC) of the dorsal visual stream integrate retinotopically encoded visual information with proprioceptive information about the position of gaze relative to the head, of the head relative to the trunk, and so on, to effect coordinate transformations that are necessary for making appropriate visually guided actions with various parts of the body.
- The transformation from an egocentric to an allocentric (i.e., environment-centered) reference frame is observed in the activity of place cells in the hippocampus.
- The metaphor of attention as a spotlight highlights many of its properties: increasing the salience of one (or a few) among many stimuli; the distinction between attending to a location vs. attending to an object; the distinction between the *site* of action of attention vs. its *source*.
- ERP studies have isolated the timing, and electrophysiological signature, of the attentional selection of a visual feature.
- Extracellular electrophysiological and fMRI studies have established that spatial

attention works, in part, by elevating baseline activity in the neural pathways that process information about stimuli occupying particular locations.

- An important function of attention may be to bias the neural competition for representation that results from the simultaneous perception of multiple stimuli in the environment.

CONTENTS

KEY THEMES

UNILATERAL NEGLECT: A FERTILE SOURCE OF MODELS OF SPATIAL COGNITION AND ATTENTION

Unilateral neglect: a clinicoanatomical primer

Hypotheses arising from clinical observations of neglect

THE FUNCTIONAL ANATOMY OF THE DORSAL STREAM

Mapping *what* vs. *where* in humans with positron emission tomography

Detecting spatial maps with functional magnetic resonance imaging (fMRI)

Coordinate transformations to guide action with perception

Extracellular electrophysiological studies of monkeys

Microstimulation of PPC produces eye movements (and more)

FROM PARIETAL SPACE TO MEDIAL-TEMPORAL PLACE

Place cells in the hippocampus

How does place come to be represented in the hippocampus?

From place to long-term memory?

THE NEUROPHYSIOLOGY OF SENSORY ATTENTION

A day at the circus

Attending to locations vs. attending to objects

Mechanisms of spatial attention

Contradictory findings from fMRI vs. EEG?

Effects of attention on neuronal activity

Spatial attention

Object-based attention

TURNING OUR ATTENTION TO THE FUTURE

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

UNILATERAL NEGLECT: A FERTILE SOURCE OF MODELS OF SPATIAL COGNITION AND ATTENTION

Previous chapters have emphasized the importance of lesions, or other factors that perturb normal brain structure or function (e.g., electrical stimulation, optogenetic manipulation), as a tool for drawing inferences about the necessity of a brain region (or physiological phenomenon) for a particular aspect of behavior or function. In effect, the emphasis has been on the importance of perturbation methods for *hypothesis testing*. We begin this chapter by highlighting a different role that neuropsychology can play in cognitive neuroscience: brain lesions can be a fertile source for *hypothesis generation*. Stated another way, careful observation of the deficits experienced by patients with brain damage can suggest organizational principles that were not evident to introspection or observation of normal behavior. With particular relevance to our present focus, the neurological syndrome of unilateral neglect has raised many intriguing questions about the fundamental workings of spatial cognition and attention.

Unilateral neglect: a clinicoanatomical primer

Unilateral neglect (sometimes referred to as “unilateral spatial neglect,” or “hemispatial neglect”) is a neurological syndrome in which damage in one hemisphere, typically due to a stroke that affects the right hemisphere, results in a pronounced unawareness of information from the side opposite the lesion (typically, the left side; *Figure 6.1*). The syndrome is highly heterogeneous, in that it can vary considerably from patient to patient. Nonetheless, we have to start somewhere, and we’ll do so with a patient whose case is representative of many described in the literature. The patient, described by neurologist and cognitive neuroscientist Anjan Chatterjee (2003), suffered a stroke when occlusion of the right middle cerebral artery deprived portions of the posterior inferior parietal lobule (Brodmann areas (BA) 39 and 40) and posterior superior temporal gyrus (BA 22) of the right

hemisphere of arterial blood (*Figure 6.2*). This patient’s consequent neglect of the left side of her world was dramatic: when her left hand was held in front of her face, she opined that she was looking at the hand of the examining doctor; she ate only the food that was located on the right side of her plate; unswallowed food sometimes collected in the left side of her mouth; despite the disproportionate weakness of the left side of her body, when asked about why she was in the hospital she would refer to weakness in her right arm. Thus, her neglect of “the left” applied to both her own body and to stimuli in her immediate environment (e.g., the plate immediately in front of her).

Unilateral neglect can result from damage to many brain areas, including the temporoparietal junction, the frontal eye fields, and the caudate nucleus. Although there remains some controversy about precisely which areas must be damaged to produce neglect, *Figure 6.3* illustrates the results of one large group study.

Hypotheses arising from clinical observations of neglect

Countless clinical observations such as those summarized in *Figure 6.1* have given rise to many influential ideas about spatial cognition and attention. Perhaps the most obvious is that humans maintain an *egocentric* reference frame that is defined with respect to the midline of our body, with the right hemisphere responsible for representation of the left side of our body (*Tip 6.1*). Underlying this undisputable observation, however, is the more vexing question of *what*

Tip 6.1

The fact that unilateral neglect of the left side of space occurs much more often after right hemisphere damage than does unilateral neglect of the right side of space after left hemisphere damage has led to the supposition, although without much additional rationale, that right hemisphere structures must allocate attention across both sides of the midline, whereas left hemisphere structures do so only for contralateral body and space.

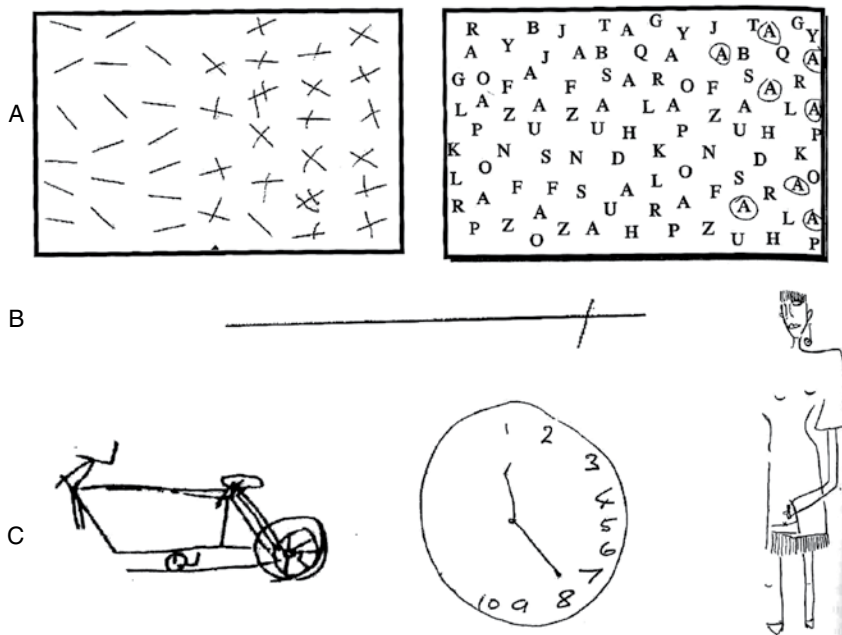


FIGURE 6.1 Patient performance on classic neuropsychological tests of neglect. **A.** One patient's performance is superior on a simple cancellation task that only requires marking each stimulus vs. on a more difficult task that requires identifying and circling targets ("A"s) among distractors (other letters). **B.** Line bisection, in which subject was asked to mark the center point of the horizontal line. Note that for "bedside paper-and-pencil" tasks such as these, direction of gaze, orientation of body, etc., is typically not controlled. **C.** Drawing from memory. Note that neglect is evident at multiple scales, in that not only are the left side of the overall pictures of the bicycle and woman neglected, but so too are the leftward-facing handlebar and leftward-facing pedal of the bicycle, and the left side of the handbag strap and left-sided earring of the woman. (Note that "left" is defined here in retinotopic terms, in that, e.g., the neglected handlebar and earring are both on the right in object-centered terms.) Source: Driver, Vuilleumier, and Husain, 2004. Reproduced with permission of The MIT Press.

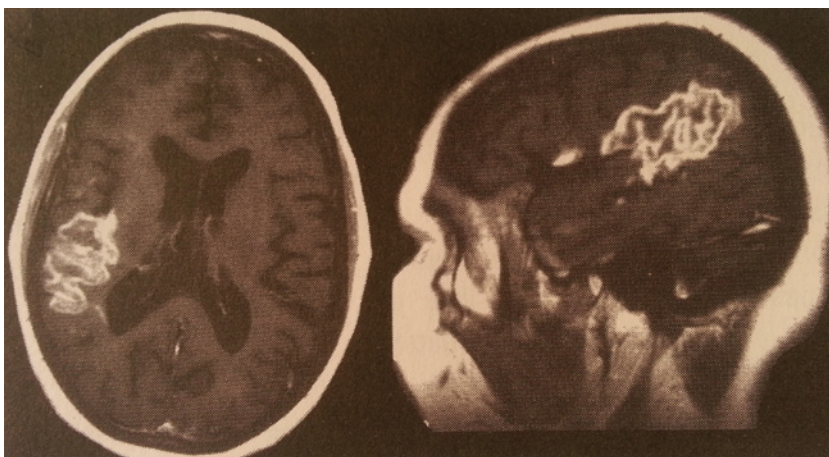


FIGURE 6.2 MR scan of unilateral neglect patient described by Chatterjee (2003), with bright regions corresponding to the damaged tissue in right inferior parietal lobule and superior temporal gyrus. (Note that radiological convention is to display the right side of the brain on the left side of the image.) Source: Chatterjee, 2003. Reproduced with permission of The MIT Press.

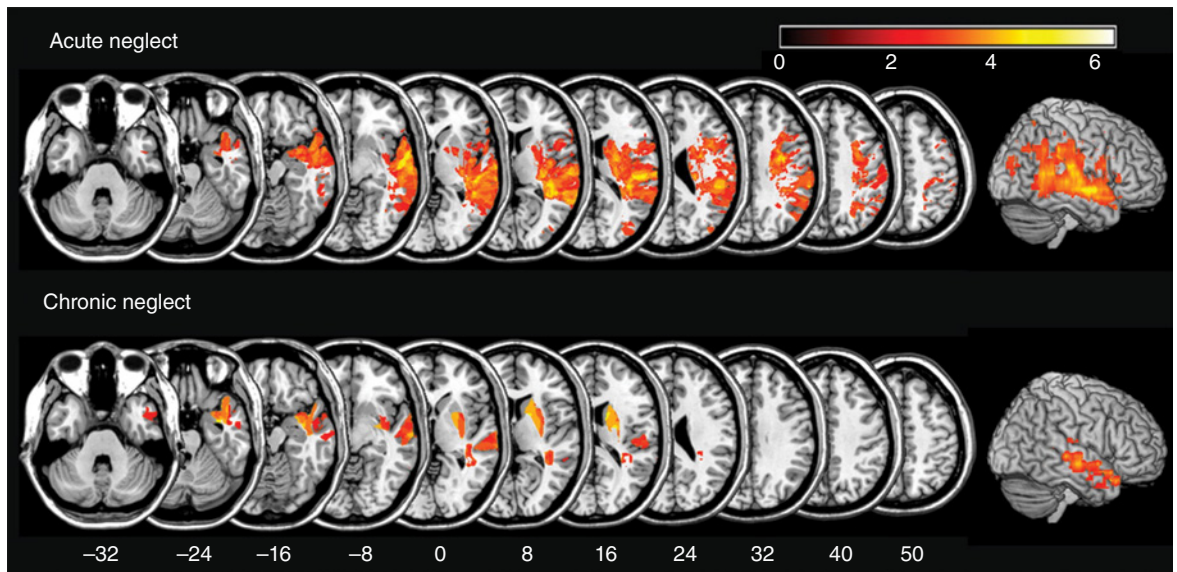


FIGURE 6.3 The neuroanatomy of unilateral neglect. This figure illustrates statistical maps of the regions most likely to be damaged in patients suffering from acute neglect (as assessed an average of 12.4 days after stroke onset) and chronic neglect (as assessed an average of 490.8 days after stroke onset). The prominence of subcortical regions in these analyses suggests important roles in spatial attention for not only gray-matter structures like the basal ganglia, but also white-matter connections between cortical regions, in particular, the inferior occipitofrontal fasciculus. The results displayed here were generated by correlating lesion location (as measured on MR scans, such as shown in Figure 6.2) with neglect symptoms in a sample of 54 patients with right hemisphere stroke. (This approach, known as “voxel-based lesion-symptom mapping (VLSM),” will be detailed in *Methodology Box 18.1*.) In this sample of 54, 24 patients had neglect symptoms during the acute stage, and a remaining 8 still exhibited neglect symptoms when reassessed in the chronic stage. Color coding indicates the t -value of the lesion-to-deficit correlation; numbers under the slices in the “chronic neglect” row indicate the distance (in mm) of each slice from the conventionally defined origin, along the dorsoventral (a.k.a. “z”) axis. Source: Karnath, Rennig, Johannsen, and Rorden, 2011. Reproduced with permission of Oxford University Press.

mechanism(s) underlie this phenomenon. Perhaps the most intuitive account is to appeal to attention, and the possibility that attention to our bodies and immediately adjacent areas is handled by posterior contralateral cortex. Although attention has many definitions, here it refers to our ability to prioritize one stream of afferent sensory information over others. (In Figure 4.6, for example, we considered the effects on the auditory evoked response (AEP) of attending to visual information from a book while clicks were simultaneously presented auditorily, versus attending to the clicks.) An alternative to attention, however, is suggested by noting that all of the examples referred to thus far – neglect of one’s own hand, of food before one on a plate (or in one’s mouth!), of

the left side of lines on a piece of paper – can be portrayed as involving the motor system. Such observations have led to the postulation that the egocentric reference frame may arise from the *intention* to perform an action either with or on a part of our body. Thus, a second theme that will recur in this chapter is the debate over whether the principle of *intention* or *attention* better captures the functions of the parietal lobes. (Relatedly, in subsequent chapters we shall consider the theoretical proposition that attention is a consequence of the intention to act – the “premotor theory” of attention.)

The *intention vs. attention* interpretation of unilateral neglect was the explicit focus of a highly influential study by the Italian neuropsychologists Edoardo

Bisiach and Claudio Luzzatti, who asked two stroke patients, both lifelong residents of Milan, to describe their city's most famous outdoor plaza, the Piazza del Duomo. Note that these sessions took place in the hospital, and, therefore, the patients' descriptions were generated not from visual perception, but from memory. A simple, but breathtakingly insightful, manipulation that they included was to ask their patients to describe the plaza from two perspectives: first, while "facing" the cathedral from the opposite side of the plaza; and second, while "standing" on the steps of the cathedral and "looking" across the plaza at the location where they had previously been "standing." Both patients, while imagining facing the cathedral from across the plaza, named several landmarks located on its south side, which would be on one's right when standing at that location and facing east. Only one of the two patients also named three landmarks on the north side (he had named nine on the south). When imagining standing on the steps of the cathedral, and therefore imagining facing in the opposite direction (i.e., facing west), in contrast, the patients each spontaneously named several landmarks on the north side of the square (i.e., on what was now their "right"), and neglected to name any of those on the south side, despite the fact that they had named several of these just moments before (*Figure 6.4*)! Thus, it wasn't that their mental representations (i.e., their memories) had been somehow cleaved in two by their strokes. Rather, it was their ability to direct their attention to (i.e., to think about) the left side of the representation as they were mentally interrogating it at that moment. From these findings, Bisiach and Luzzatti (1978) concluded that "unilateral neglect cannot be reduced to a disorder confined to the input-output machinery of the organism interacting with its physical environment" (p. 132). That is, they argued that neglect is a disorder of attention, not intention.

In their writings, Bisiach and Luzzatti (1978) emphasized that neglect revealed a deficit in mental representation. Certainly, when the performance of their patients is compared with that of the patient illustrated in *Figure 6.1*, one sees that patients can neglect the left side of individual objects, even when their drawings suggest that they are aware that other objects and/or parts of the environment exist to the left of the object in question. This provides at least a

hint for the existence of an *object-based attention* that operates at the level of individual objects, rather than large swaths of the visual field.

Yet a different concept – that we may represent "near" vs. "far" space with different mechanisms – is illustrated by the case of a patient who very well may not have produced the same results as those reported by Bisiach and Luzzatti in 1978. This patient was a British mechanic, tested in a stroke rehabilitation center in Oxford, UK. The research team first tested the patient on line bisection using the standard method: a horizontal line drawn on a piece of paper was held 45 cm from his eyes, centered on his midline, and he was asked to mark the midline with a pen. The results (from several trials, with lines of varying length) were comparable to that shown in *Figure 6.1*, demonstrating marked neglect of the left side of the line. Next, the patient was tested with lines placed on a whiteboard that was located 2.44 m distant from the patient (a distance chosen because it corresponded to the standard distance between dart board and dart thrower that one finds in a British pub). The thickness and length of the lines were scaled up, such that they subtended roughly the same amount of visual angle as had the lines on paper from the "45 cm" test. Instead of drawing with a pen, the patient was asked to indicate the midpoint of the across-the-room lines by pointing with a light pointer (equivalent to a laser pointer, but this was 1990) or, in a cheeky follow-up test, by throwing darts at it. Whether with light pointer or darts, the patient showed no systematic bias away from the midpoint: his errors clustered near the true midpoint, and half of them fell to the left of it. Thus, this case demonstrated concretely a dissociation between the processing of **peripersonal space** – the space surrounding one's body that is within one's grasp – vs. **extrapersonal space** – those portions of perceived space that are further than we can reach.

A final concept that we will consider here is that studies of neglect also provide compelling evidence for *competition* within the brain. **Extinction**, in the neurology clinic, refers to the fact that a patient can sometimes be aware of a single object presented on her left, but will lose awareness of it if a second item is then presented on her right. It's as though the item on the right extinguishes awareness of the item on the left.

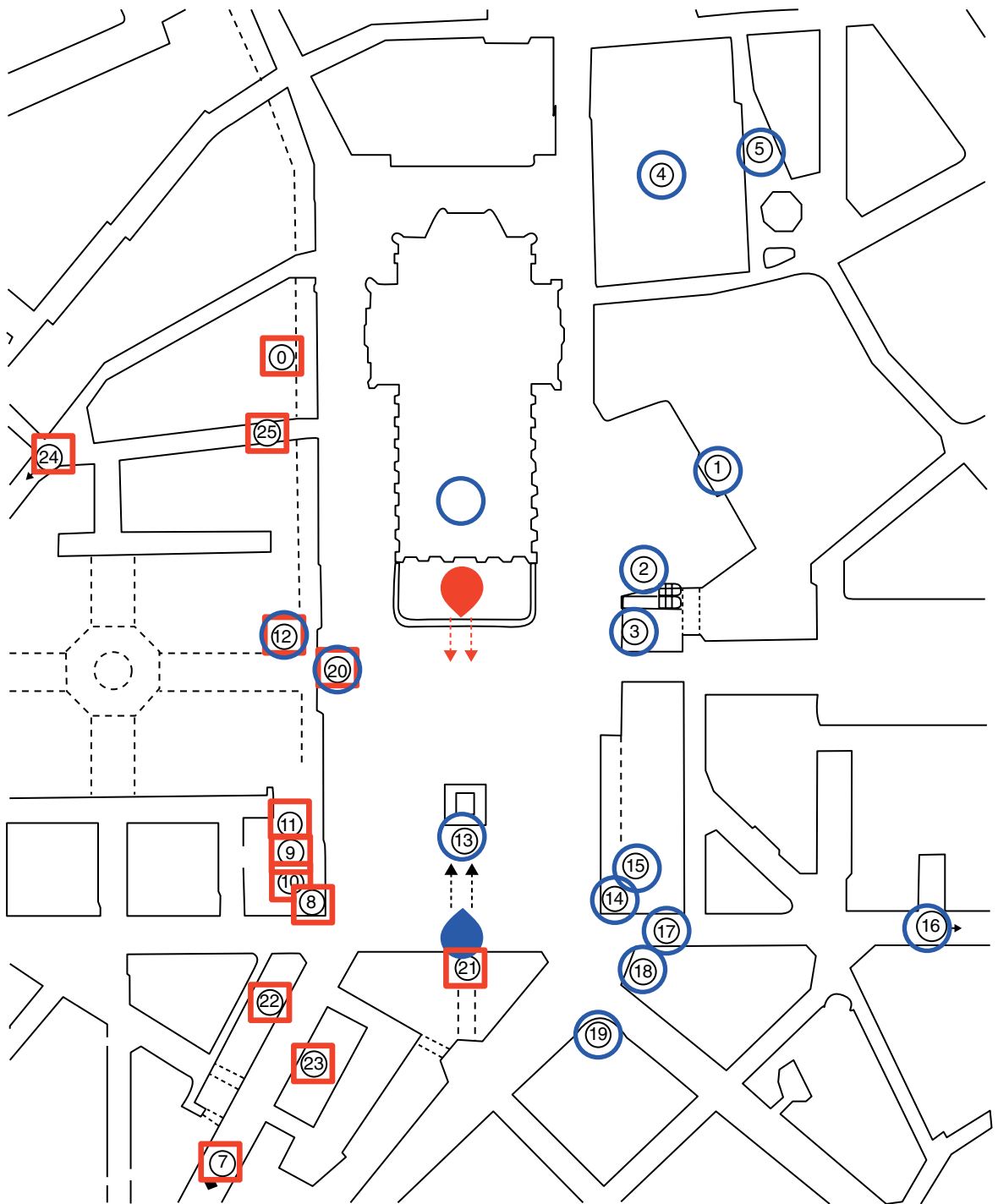


FIGURE 6.4 Landmarks named by two unilateral neglect patients of Bisiach and Luzzatti (1978). When imagining themselves looking at the cathedral from across the plaza (i.e., facing east, as cartooned by blue head and two arrows), they named the landmarks highlighted with blue circles; and when imagining themselves standing on the cathedral steps and looking in the opposite direction (i.e., facing west, red head and two arrows) they named the landmarks highlighted with red squares. Source: Adapted from Bisiach and Luzzatti, 1978. Reproduced with permission of Elsevier.

Each of these ideas about how the brain may represent space and allocate attention will be revisited in this chapter as we consider the anatomy and physiology of spatial processing and attention in neurologically intact organisms (be they humans, monkeys, or rats). Some examples of behavior characteristic of unilateral neglect are viewable at *Web Link 6.1* and *Web Link 6.2*.

THE FUNCTIONAL ANATOMY OF THE DORSAL STREAM

Mapping what vs. where in humans with positron emission tomography

As a result of technical advances in medical imaging in the 1970s and 1980s, by the mid-1980s it became possible to begin to perform neuroimaging studies on neurologically healthy human subjects. The first technology to become available for cognitive neuroscience research was positron emission tomography (PET). As illustrated in *Methodology Box 6.1* and *Figure 6.5*, the first application of the new technology to human cognition was in the context of that uniquely human trait, language. It didn't take long thereafter, however, until PET scanning was also being applied to visual perception. In one relatively early, and influential, study of the visual system, James Haxby and colleagues at the National Institutes of Health used PET to demonstrate a "Dissociation of object and spatial visual processing pathways in human extrastriate cortex" (*Research Spotlight 6.1* and *Figure 6.6*). In historical perspective, the significance of this study was due less to any genuinely new insights about brain function that it provided, and more to its contribution to a conceptual buttress on which much of the cognitive neuroscience enterprise has relied from that time forward: it demonstrated a clear linkage between what had been inferred about human brain function from neuropsychological data, such as from the first part of this chapter, and neurophysiological data acquired primarily from nonhuman primate proxies for the human brain. Further along in this chapter we'll go into more detail about we've learned from the latter, but for now, let's recall from *Figure 5.3* that a characteristic of receptive fields of neurons in the dorsal stream of the monkey is that

they typically represent contralateral regions of space. Do analogous properties exist in the human dorsal stream?

The methods featured in *Research Spotlight 6.1* did not have the temporal resolution to address the question about receptive fields in the human dorsal stream. This is because of an inherent limitation in the temporal resolution of the $H_2^{15}O$ PET method. First, it takes some time for the tracer, once injected into the bloodstream, to make it into the organ whose activity one wants to measure. Second, the sensitivity of the method is such that several minutes of data acquisition are required to be able to reliably detect differences between, say, performance of dot-location matching vs. performance of a sensorimotor control task. *Figure 6.6.B* illustrates this point nicely. Let's take "viewing words" vs. "visual fixation," for example, and focus on the region near the temporo-occipital junction where the "averaged difference [i.e., subtraction] image" shows a splotch of yellow and red, indicating greater rCBF for words vs. fixation. The color of this region in the "Averaged blood flow" images suggests values of roughly 1000 PET counts (just an estimate from eyeballing the color scale). That means that the signal emanating from that region increased from ~1000 PET counts, measured from four minutes of fixation, to ~1040 PET counts measured from four minutes of word viewing – a 4% change. (And these are just "relative PET counts"; the Haxby et al. (1991) paper reports that a range of 1.7–4.1 million counts (i.e., photon collisions with the scanner's detectors) were measured in each slice across each four-minute scan.) Now, back to dot-location matching. Within each four-minute scan, a subject might perform, let's say, 40 trials. On each trial the dots need to appear in different locations, of course, to ensure that the subject is engaging her spatial perception machinery at a steady level throughout the scan. At the end of the scan, the experimenter gets one number (e.g., 1040 relative PET counts) corresponding to the average level of rCBF in a (**region of interest (ROI)**) in, for example, parietal cortex. There is simply no way to deduce from that one number whether that ROI may have been more active, for example, on individual trials when the dot was located in the subject's right visual field vs. when the dot was located in the left visual field. To obtain this trial-level resolution, one

METHODOLOGY BOX

6.1 X-ray computed tomography (CT) and positron emission tomography (PET)

The word tomography refers to the acquisition of an image of a 2-D section through a volume, via some kind of penetrating wave. Although X-ray technology was used for medical imaging throughout the twentieth century, it was limited to the acquisition of single 2-D sections (or “slices”) until the advent of computer-implemented algorithms that allowed for the effectively simultaneous collection of all the slices needed to reconstruct a volume (from *Figure 6.3*, one can see how stacking all the 2-D slices yields a 3-D volume). (Note that, although these sections are referred to as being “2-D slices,” each actually has some thickness to it, and thus they might be better construed as thin “slabs.”) CT is limited to acquiring structural images (*Tip 6.2*).

PET entails the injection of a short-lived (on the order of minutes to hours) radioactive isotope into the blood stream, and subsequently measuring where in the body this substance accumulates. It works on the principle that the decay of the isotope releases sub-atomic particles called positrons, and that when a positron encounters an electron the resultant annihilation releases two photons (“gamma rays”) that travel in opposite directions. The photons are detected by detectors that are arrayed in a ring around the body, and sophisticated math is performed to estimate where in space the annihilation event must have occurred in order to produce the pattern of photons detected by the scanner.

In principle, any biologically active molecule that can be rendered radioactive can be used as a PET “tracer.” Thus, PET can be used to assess patterns of energy metabolism in the brain (via radiolabeled glucose, a method that can be thought of as “in vivo 2DG imaging” (*Methodology Box 3.1*), or patterns of activity of certain neurotransmitter systems (as with, e.g., radiolabeled dopamine agonists). For now, we’ll limit our discussion to H_2^{15}O – a.k.a., “radiolabeled water.” In a “cognitive activation study” (such as that profiled in *Research Spotlight 6.1*), the subject lies in the scanner, with an IV in her arm. Upon the delivery of H_2^{15}O into the blood stream, she begins performing a cognitive task. At this point, it is useful to think back to *Thought Box 2.2: The metabolic demands of the action potential and synaptic activity*. In a face-matching task, for example, increases in the activity of the brain tissue engaged in processing this visual information will trigger an increase in the delivery of arterial blood to this region. (The physiology of this process remains poorly understood, although it may be that elevated cellular activity releases a gas, such as nitric oxide (NO), whose diffusion causes the dilation of capillaries in the vicinity.) With H_2^{15}O PET, such changes in regional cerebral blood flow (rCBF) are quantified as changes in “PET counts” detected by the scanner (i.e., a high number of photons emanating from a location in the brain indicates that there is a high concentration of H_2^{15}O molecules at that location).

Because the entire brain is always consuming large amounts of arterial blood (20% of the energy consumption of the body), task-based imaging requires image subtraction in order for the results from the scanning session to be interpretable (*Figure 6.5*).

would need a method whose signal-to-noise ratio (SNR) is high enough to produce data that are interpretable at second-level resolution, as opposed to the

Tip 6.2

Although the quality and resolution of CT images are markedly inferior to those produced by MRI, CT is nonetheless still commonly used at hospitals. This is particularly true for emergency rooms, because it presents no risks to a patient whose medical history is unknown. (E.g., it could be catastrophic to put a comatose patient into an MR scanner if that individual had metal aneurism clips, or a pacemaker.)

minute(s)-level resolution provided by PET. Indeed, at the same time that Haxby and colleagues were publishing their 1991 study, a method was being developed with just such characteristics: *functional magnetic imaging (fMRI)*.

Detecting spatial maps with functional magnetic resonance imaging (fMRI)

Recall from *Methodology Box 2.1*, about MRI, that *different pulse sequences can emphasize different types of tissue, and different ways of measuring the relaxation signal can yield different kinds of information about the tissue being imaged*. As outlined in *Methodology Box 6.2*, research in

FIGURE 6.5 PET scans of word processing, illustrating the logic of image subtraction.

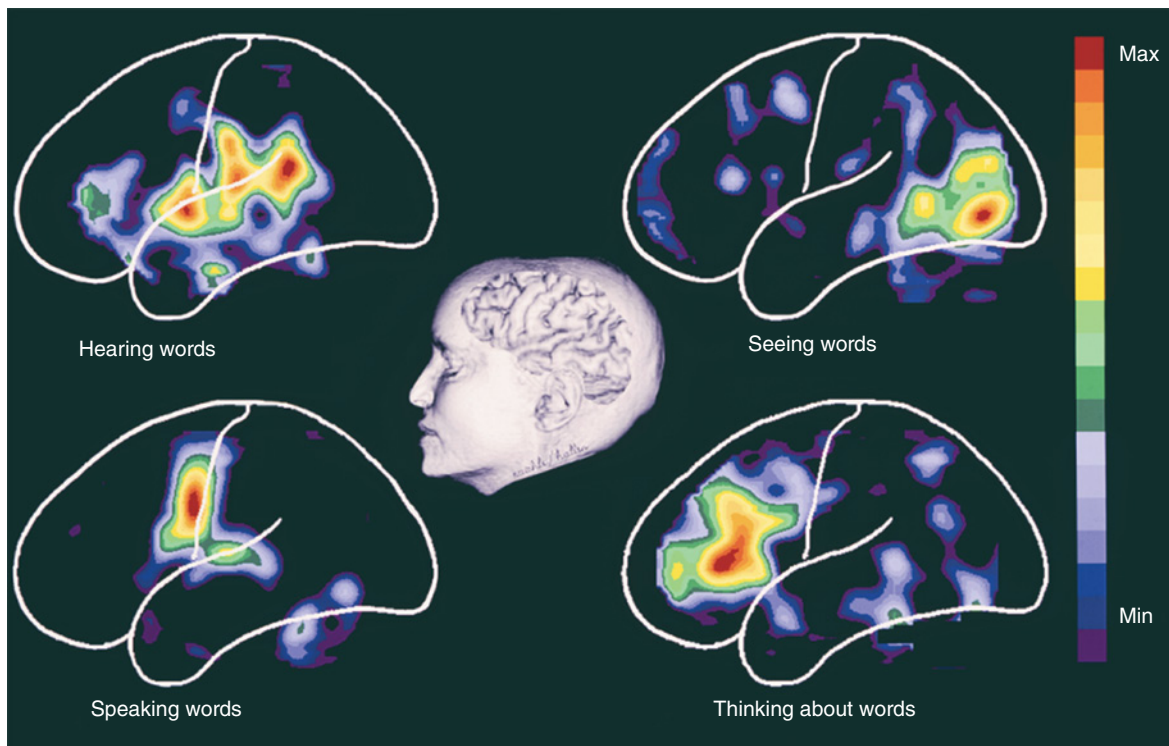


FIGURE 6.5.A PET subtraction images of rCBF when subjects perform the labeled cognitive operation versus a matched control condition. For *Hearing words* and *Seeing words*, the contrast is versus fixation of a point on the screen. For *Speaking words*, the contrast is versus *Seeing words*. For *Thinking about words*, the task was to think about a use for the presented word (e.g., for the word “cake,” the subject might think “eat”), and the contrast is versus *Speaking words*. Source: Petersen, Fox, Posner, Mintun, and Raichle, 1988. Reproduced with permission of Nature Publishing Group.

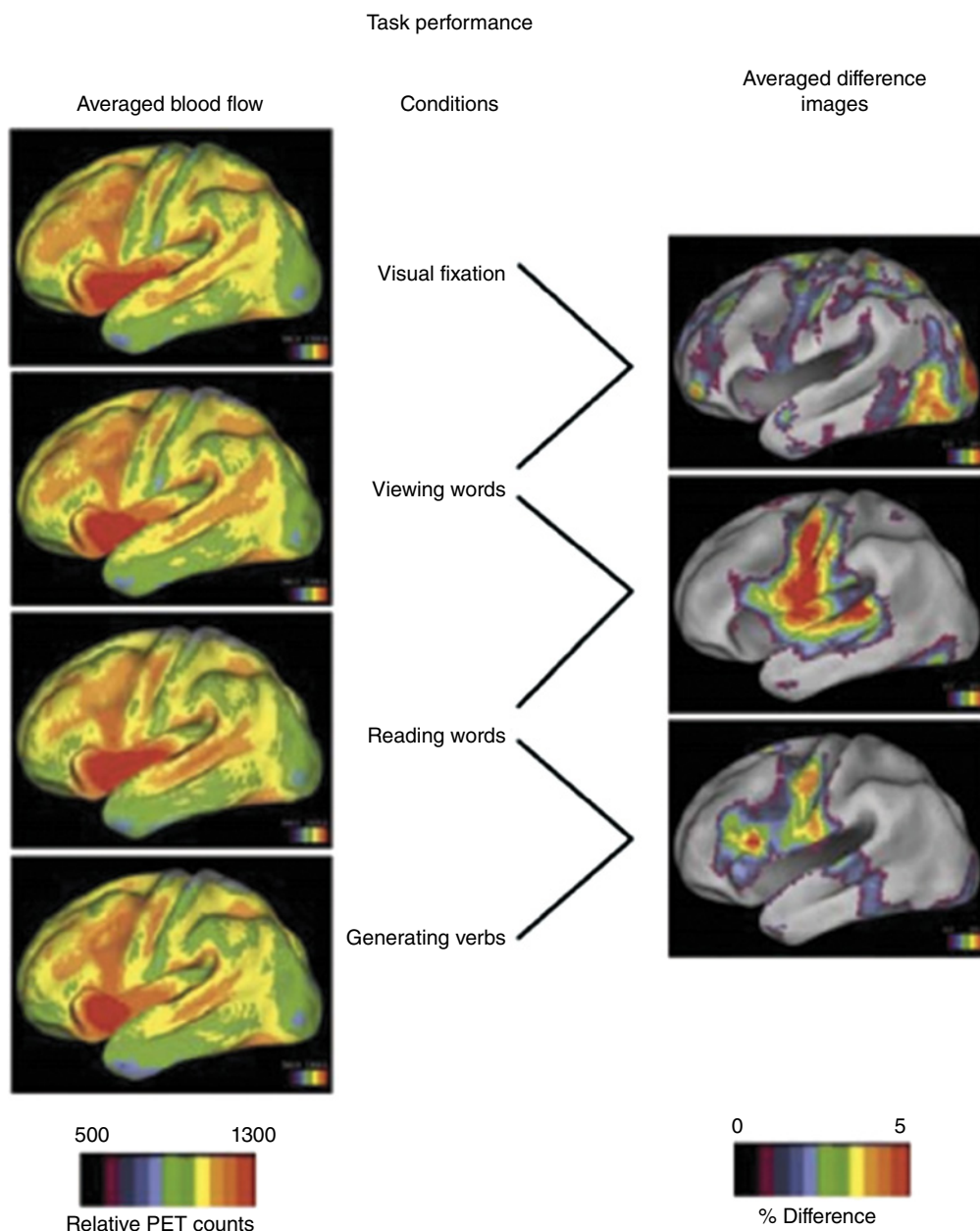


FIGURE 6.5.B A re-representation of the Petersen et al. (1988) data. The right-hand column presents three of the rCBF contrast images from Figure 6.5.A (now looking more “modern” because the activity maps are warped onto inflated left hemispheres). The top image is *Seeing Words*, the middle image is *Speaking Words*, the bottom *Thinking About Words*. Note that no task-related changes in rCBF exceed 5% of the baseline level of rCBF. The left-hand column shows the unsubtracted rCBF images associated with each of the four conditions used to generate the subtraction images in the right-hand column. Try covering up the labels, as Dr. Raichle is wont to do during presentations, and see if you think you’d be able to identify which unsubtracted rCBF image corresponds to which behavioral condition. The point that he is making hearkens back to Chapter 2’s point that much of the brain’s “churning” seems to have little to do with moment-to-moment stimulus processing and behavior. Source: Raichle, 2010. Reproduced with permission of Elsevier.

RESEARCH SPOTLIGHT

6.1 Exploring what vs. where in the human visual system, with PET (Haxby et al., 1991)

One of the first generation of cognitive neuroimaging studies, this study exemplifies both “best practices” for the methods available at the time (many of which still apply today), as well as some omissions that wouldn’t pass the scrutiny of peer review in this day and age. The behavioral task had strong construct validity in that it was adapted from the well-established and validated Benton Facial Recognition Test. It was balanced with regard to task demands in that subjects had to follow the same procedure in both conditions: in the face condition, subjects indicated which of the two faces in the lower boxes matched the face in the upper box; in the dot-location condition, they indicated which of the two dots in the lower boxes was in the same location relative to the double line as the dot in the upper box. Finally, subjects also performed a control condition – pressing a button each time three empty squares were presented – that included the same sensory and motoric elements as the experimental tasks, but not the elements of central interest – face processing and spatial processing.

Three PET scans were acquired per subject, measuring task-related changes in rCBF via H_2^{15}O (*Methodology Box 6.1*) while subjects performed each of the three tasks. Each scan lasted four minutes, the subject performing several trials of the task during each scan. (Omitted in the published report is information about the rate at which trials were presented, as well as information about how accurately subjects performed on the task. Why might these factors be important for interpreting the neuroimaging data?) The data were analyzed in two steps. First, face-matching maps and dot-location matching maps were computed for each subject by subtracting each of the experimental scans from the control scan. *Figure 6.6* illustrates the regions at which two or more subjects showed statistically greater activity during the experimental than the control task.

FIGURE 6.6 PET scanning demonstrates “Dissociation of object and spatial visual processing pathways in human extrastriate cortex.” Source: Haxby, Grady, Horwitz, Ungerleider, Mishkin, Carson, Herscovitch, Schapiro, and Rapoport, 1991. Reproduced with permission of PNAS Group.

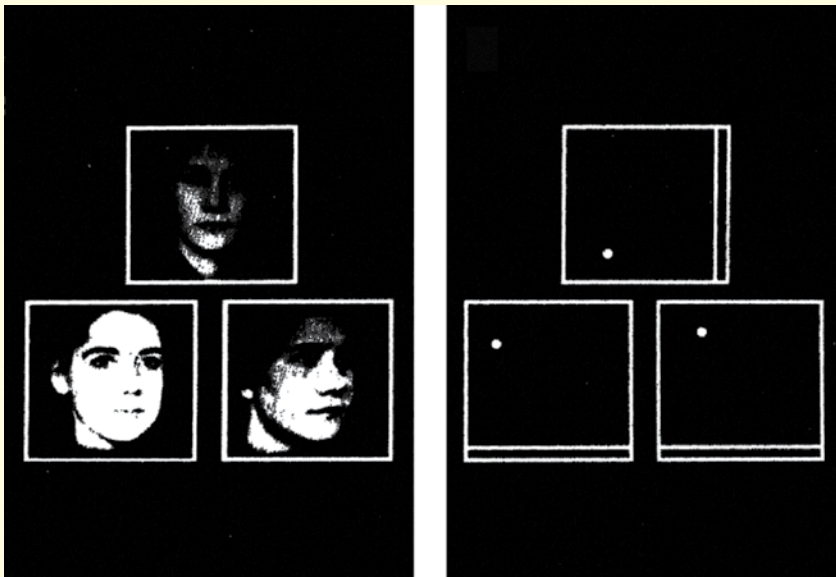


FIGURE 6.6.A Illustration of the two visual discrimination tasks.

Research Spotlight 6.1 *Continued*

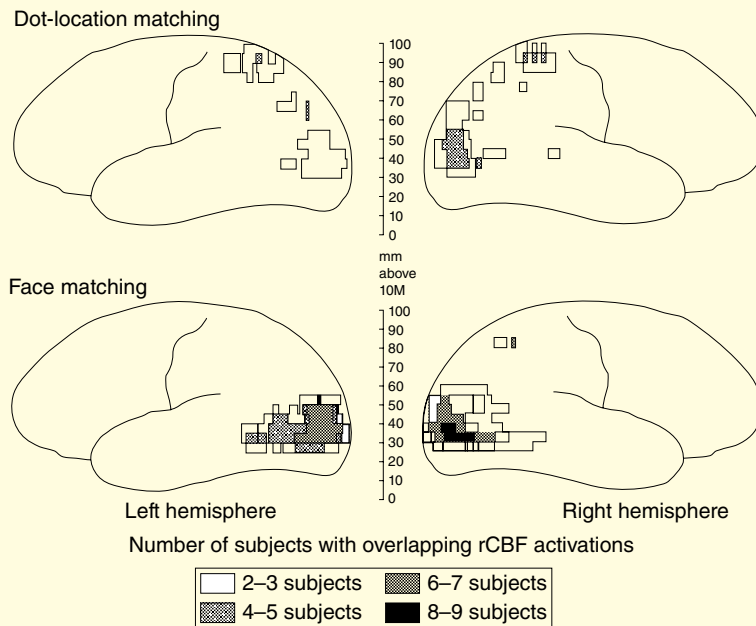


FIGURE 6.6.B Although the precise regions of elevated rCBF showed less overlap across subjects in the visuospatial task (top row), 8 of 11 subjects showed elevated rCBF in lateral occipital regions in this condition, and 10 of 11 did so in parietal regions. For the face discrimination task, 10 of 11 subjects showed elevated activity in both lateral occipital and in occipitotemporal cortex regions of interest.

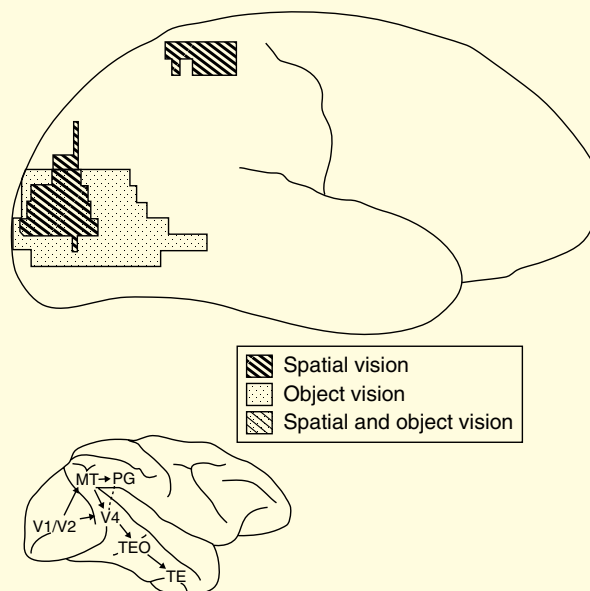


FIGURE 6.6.C Top image is a composite of the (human) PET data, with left-hemisphere data projected onto the right hemisphere, showing cortical regions where three or more subjects showed greater experimental-than-control rCBF. Bottom image is the brain of a macaque monkey, drawn to scale relative to the human brain, labeled to identify the what and where systems.

METHODOLOGY BOX

6.2 Functional magnetic resonance imaging (fMRI)

Recall from *Methodology Box 2.1* that MRI is based on perturbing the axis of rotation of protons, and measuring of the energy given off as they return to their low-energy state. For structural MRI, this energy is called the T1 signal. In 1990, Japanese biophysics researcher Seiji Ogawa and colleagues, working at AT&T Bell Laboratories in Murray Hill, NJ, made a breakthrough discovery: leveraging the fact that oxygenated hemoglobin (Hb) and deoxygenated hemoglobin (dHb) have different magnetic profiles, they were able to coax from MRI a “blood oxygen level-dependent” (BOLD) signal. That is, with the right pulse sequence they demonstrated that a different MRI measure of relaxation, called T2*, varied with the balance between Hb and dHb. The reason that this was important comes from a principle of brain physiology.

Thought Box 2.2 established the cellular factors underlying the fact that “elevated activity in neurons creates a need for arterial blood.” This results in dilation of arterioles, and an increase of arterial blood carrying Hb in the vicinity of the active cells. A distinctive feature of mammalian physiology is that this hemodynamic response is invariably overcompensatory, in that the increased supply of Hb-rich blood exceeds the demand and, as a result, the ratio of Hb:dHb in venules carrying blood away from the capillary bed temporarily increases. Neurosurgeon Wilder Penfield (of somatosensory homunculus fame) observed this with his naked eye over 70 years ago, as venous blood becoming redder during localized seizures. *Figure 6.7* illustrates the hemodynamic response to neuronal activity in a cartoon: the elevated activity of a small cluster of neurons (depicted by coloring them yellow) results in an increase in the delivery of Hb-rich blood to the local capillary bed (depicted by redder tint of capillaries), and in an increase in Hb:dHb in the venules draining this blood away from the capillary bed (also depicted by a redder tint). The way that this looks in fMRI data is illustrated in *Figure 6.8*.

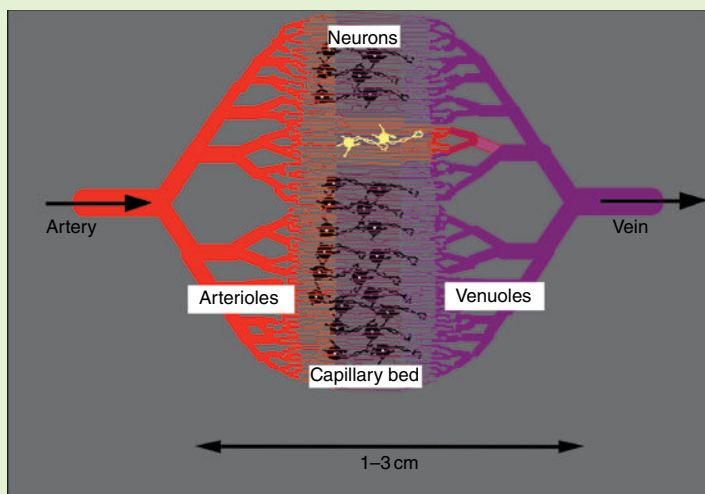


FIGURE 6.7 Cartoon of the hemodynamic changes produced by neuronal activity. Source: Menon and Kim, 1999. Reproduced with permission of Elsevier.

Methodology Box 6.2 Continued

Figure 6.8.A shows data acquired from a 3T scanner: axial slices of the MRI of a subject, with a statistical map derived from fMRI data superimposed. Each T1-weighted anatomical slice is made up of a grid of 256×256 voxels from which data were acquired. In this image, the anatomical voxels measure $0.9375 \text{ mm} \times 0.9375 \text{ mm}$ in-plane, and are 4 mm thick. The fMRI data are of coarser resolution, acquired from a 64×64 voxel grid in which each voxel measures $3.75 \text{ mm} \times 3.75 \text{ mm} \times 4 \text{ mm}$. To generate this activation map, a statistical test (a general linear model (GLM), effectively a multiple regression) was performed at each of

FIGURE 6.8 fMRI signal from a scan lasting 6 minutes and 40 seconds, during which the subject pushed a button with each thumb, once every 20 seconds (see *Web Clip 3.1* and *3.2*). Source: © Bradley R. Postle.

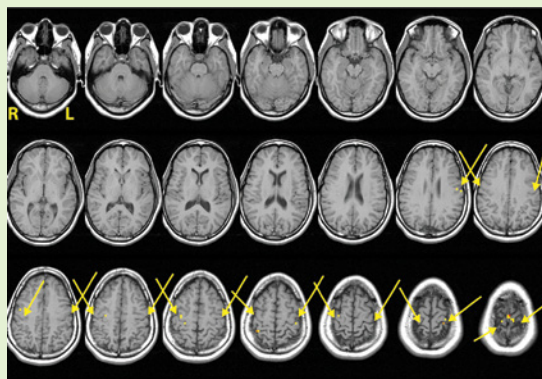


FIGURE 6.8.A Voxels in the pre- and postcentral gyrus for which the increase fMRI signal in response to the button presses exceeded a statistical threshold. This “thresholded statistical map” is superimposed on a T1-weighted anatomical image acquired during the same scanning session. This mosaic displays axial slices arrayed from ventral-most (upper left) to dorsal-most (lower right). Arrows point to the central sulcus in each slice in which it is visible. (Primary motor cortex is located in the precentral gyrus (*Chapter 7*).)

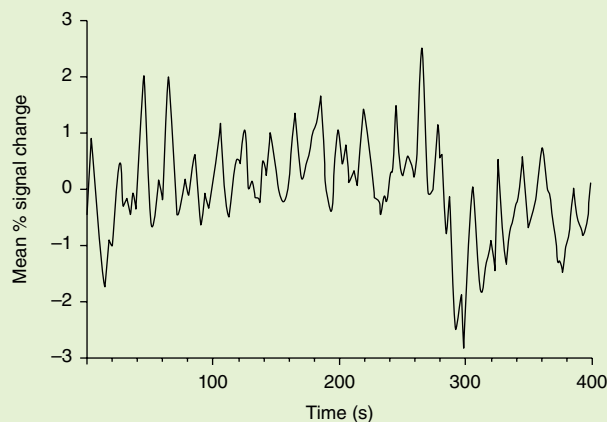


FIGURE 6.8.B BOLD signal fluctuations pooled across the voxels identified in Figure 6.8.A. The data have been mean centered and scaled to units of percentage change from baseline, but are otherwise “raw.” A button press occurred at each of the tick marks on the time axis (except 400).

Methodology Box 6.2 *Continued*

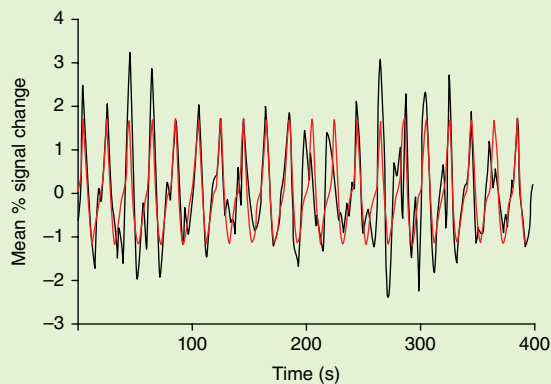


FIGURE 6.8.C Same data as *Figure 6.9.B*, but with drift removed, and an overlay (red waveform) of the fits of the statistical model to the data (i.e., the general linear model). With regard to “drift,” the overall drop in signal at time 280 in *Figure 6.8.B* reflected some factor other than the once-every-20-seconds button press, and so was removed statistically. The independent variable in the statistical model was a series of delta functions (a.k.a. “stick functions”) placed at time 0, 20, and so on up to time 380 seconds – i.e., at each time point when a button press was to be executed – and convolved with a basis set of sine waves and cosine waves, so as to capture the predicted shape of the HR. Whereas the red waveform shows the least-squares best fit of the model to the data, the data themselves illustrate the trial-to-trial variability of the fMRI response to each of the 20 presumably identical events.

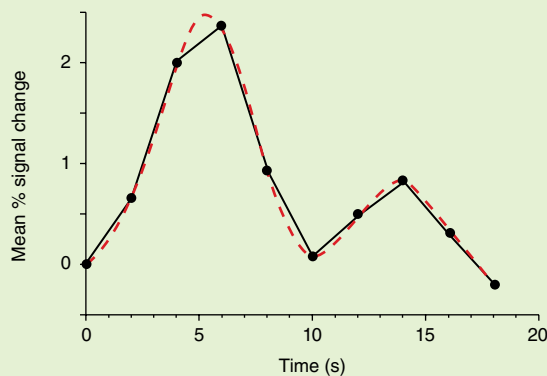


FIGURE 6.8.D In black, the trial-averaged response to a bimanual thumb press made at time 0 (i.e., the average of time segments 0–20, 20–40, ... 380–400 from *Figure 6.8.C*). This is an estimate of the hemodynamic response function (HRF) for this subject. In red, the HRF has been interpolated, to represent the fact that it is a continuous function, despite only being sampled at a 0.5 Hz rate in this study.

the roughly 30,000 voxels in this subject's brain, evaluating at each one evidence for systematic signal change corresponding to the subject's once-per-20-seconds behavior.

Figure 6.8.D illustrates that, although the neural activity producing the button press was probably completed by time 0.5 s or so, this subject's fMRI response didn't peak until time 6 s due to the sluggishness of the hemodynamic response. Thus, the temporal resolution of fMRI is an order of magnitude coarser than EEG/ERP, and an order of magnitude finer than PET.

See *Web Clip 3.1* and *Web Clip 3.2* for more on fMRI, *Web Link 6.3* for an intriguing application of fMRI for neurofeedback, and *Web Clip 6.1* for ways that “resting state” fMRI signal can be analyzed to study the brain's functional organization.

the 1980s and 1990s yielded pulse sequences that produced a signal that is sensitive to the proportion of oxygenated-to-deoxygenated hemoglobin in blood – the **blood oxygen level-dependent (BOLD)** signal. This enabled *functional* MRI, in that the dynamics of arterial blood contribute to the function of the brain (i.e., more of it is summoned to regions experiencing elevated cellular activity). There are two important qualities of fMRI that give it a temporal resolution that is superior to PET. First, its higher SNR means that interpretable data can be acquired from much shorter periods of time (i.e., it's not necessary to integrate over four minutes of scanning time to obtain an interpretable measure). Second, the fact that it measures an intrinsic signal, and thus doesn't entail injecting a radioactive tracer into the subject's body, means that much more data can be collected from each research subject. (The only practical limit on how long a subject can be kept in the MRI scanner is his ability to lie still and stay alert. That, and the experimenter's research budget, because fMRI scanning is typically charged at an hourly rate.)

There have been many, many important experiments using fMRI to study the neural bases of human vision. At this juncture, we'll consider just one that succinctly illustrates the two aspects of cognition that are the focus of this chapter: the representation of space, and attention. In it, Saygin and Sereno (2008) employed a variant of a method that is commonly used to perform retinotopic mapping – the “rotating wedge.” The logic is that, while the subject fixates a central location, a wedge shaped like a slice of pie is slowly rotated around the fixation point. The wedge looks different from the background – classically it might contain an alternating black-and-white checkerboard pattern presented against a gray background – and the logic is simply that areas of the brain that represent a particular region of retinotopic space will only show increases of activity when the region that they represent is occupied by the wedge. (A variant of this is demonstrated in *Web Clip 6.2*.) Saygin and Sereno employed a clever riff on this basic logic, by presenting clusters of colored dots on the entire screen, rotating the entire display around the fixation cross, and having the clusters of dots defining the wedge take on recognizable shapes (*Figure 6.9.A*). Specifically, the recognizable shapes were “point-light walker” images, created by having an actor wear a black

bodysuit fitted with small lights on the arms, legs, head, and torso, and filming him as he performed a variety of actions (walking, jumping rope, etc.). Their rationale was that many areas of the brain outside low-level visual cortex are more likely to process information about “biologically relevant” stimuli than something as mundane as a checkerboard pattern.

Before we consider the implications of the Saygin and Sereno (2008) results for spatial cognition and attention, let's consider the fact that, whereas it's obvious how their experiment systematically stimulated different locations in the visual field, thereby allowing them to identify retinotopically selective parts of the visual system, there was nothing in the experiment that explicitly isolated each of the regions identified in *Figure 6.9*. That is, their display didn't somehow only activate V1, then V3, then IPS2, and so on. So how were they able to discriminate these regions and determine their boundaries? The answer is that they used their knowledge of the physiology and anatomy of the human visual system, including the fact that retinotopic maps “reverse” at each interareal boundary. This is particularly evident along the strip of tissue extending from V3A to V7 to IPS1 to IPS2 in the right hemisphere *Attention* image. Moving in a caudoventral-to-rostr dorsolateral direction, V3A's retinotopy transitions from horizontal to upper visual field, V7's from upper to horizontal, IPS1's from horizontal to upper, and IPS2's from upper to horizontal to lower. Known anatomical landmarks were also used to determine precisely where to draw boundaries.

Figure 6.9 reveals several important facts about the functional organization of the human dorsal stream, and we can consider these in relation to the hypotheses generated by unilateral neglect syndrome, as considered at the beginning of this chapter:

- *An egocentric reference frame.* By definition, a region whose response is sensitive to the retinotopic location of a stimulus is processing information in terms of its location relative to the body – the retinae being parts of the body ;-). Thus, many regions of the dorsal, and ventral, streams could, in principle, support the egocentric processing that is disrupted in unilateral neglect.
- *Attention vs. intention.* This study doesn't address this question explicitly, because there was no condition in which subjects were required to

generate a movement toward a stimulus (i.e., they made the same button presses regardless of where on the screen a to-be-detected event occurred.) However, it does demonstrate very clearly two consequences of the allocation of attention to a stimulus at a location in space vs. the unattended stimulation of that same retinotopic location. The first is a decrease of stimulus-evoked activity in V1 (despite the fact that the physical stimulation was

identical. Why might this be? Hint: final section of *Chapter 5*). The second is a commensurate increase in stimulus-evoked activity in many extrastriate regions, including the MT complex, the IPS, the ventral surface of the temporal lobes, and the frontal eye fields (FEF). (Much of the focus for the remainder of this chapter will be on the IPS; the FEF will get a starring role in *Chapter 8*, and the ventral temporal cortex in *Chapter 9*.)

FIGURE 6.9 Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. Source: Saygin and Sereno, 2008. Reproduced with permission of Oxford University Press.

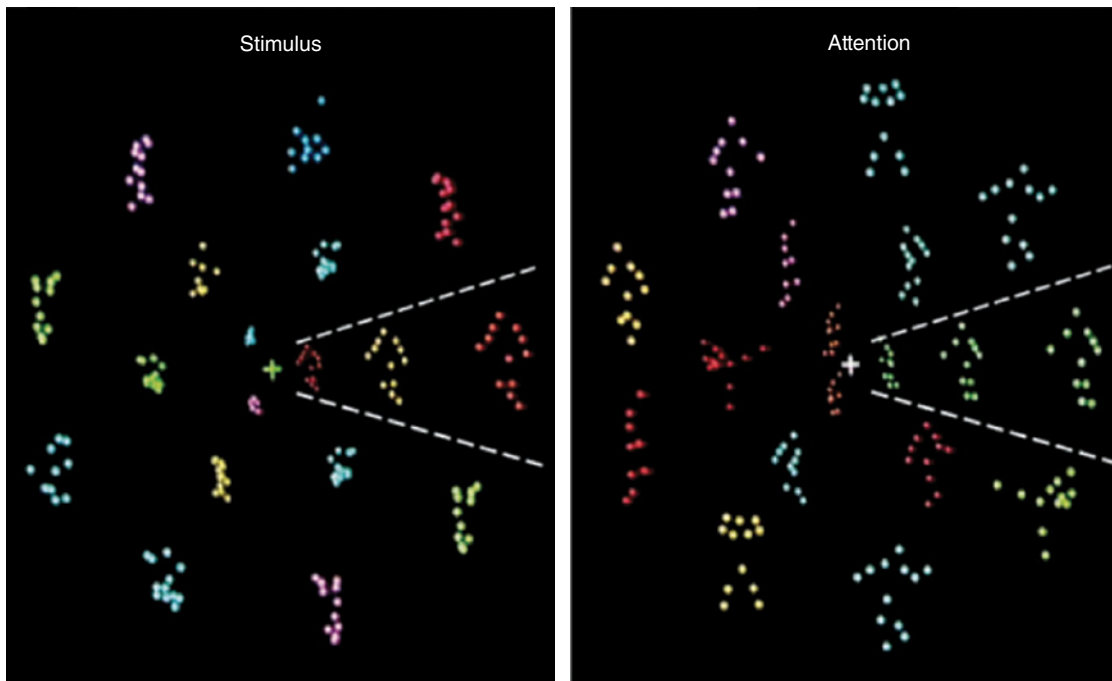


FIGURE 6.9.A A static screen shot of the stimulus displays from the two conditions. The entire display rotated about the fixation point (like a speeded-up depiction of the night sky over the course of 12 hours), with clusters continually fading in and out. In the *Stimulus* condition the wedge was defined by clusters taking on the configuration of a point-light walker (the wedge is indicated here, for clarity of exposition, with white dotted lines that were not present during the experiment; to see it as the subject did, visit *Web Clip 6.2*). In the *Attention* condition, all clusters took on the shape of point-light walkers, the to-be-attended wedge differentiated by being the only sector in which three of the same color were aligned along a single radius. In the *Stimulus* condition, the color of the fixation cross changed every second, and subjects were to report each time a change constituted a “2-back” repetition (e.g., *red-yellow-red* would be a hit, but *red-yellow-blue-red* and *red-red-yellow* would not). Note that, although the *Stimulus* condition placed heavy demands on attention, an analysis identifying regions responding to the periodic sweep of the wedge would not detect activity corresponding to steady attention to the central point on the display. The *Attention* condition, in contrast, can be construed as the attention-to-the-rotating-wedge condition, because subjects, although fixating centrally, were to respond each time the three same-color point-light walkers in the wedge were *not* all moving identically (e.g., if two were jumping rope and one was doing karate kicks).

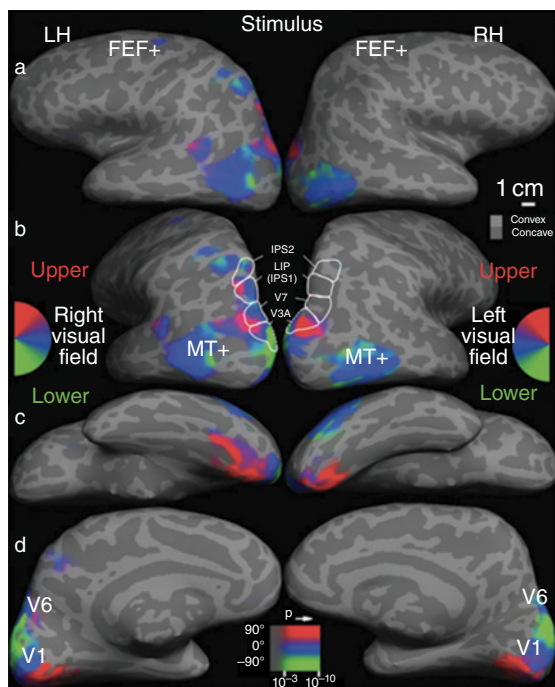


FIGURE 6.9.B Retinotopically sensitive cortex in the *Stimulus* condition, displayed on three views of an inflated hemisphere ($\times 2$, because both hemispheres are displayed); lighter gray corresponds to gyri. Superimposed on the anatomical images are patches of red-blue-green color identifying cortical regions whose activity fluctuated (either increased or decreased) each time the wedge passed through a particular region of the display, the specific color indicating the part of the visual field to which that patch of cortex was responsive (keys are located to left and right of view b for each hemisphere). Statistical significance of a region's retinotopic responses is represented by the translucency of the superimposed color. The identity of each region is labeled with white letters and, along the intraparietal sulcus (IPS), borders.

- *Object-based attention.* Although this study wasn't designed to assess the neural bases of object-based attention, the "stimulus" condition did, in fact, require object-based attention, in that it required focused attention to changes of color of the fixation cross. However, because the data were analyzed with respect to the rotating wedge, and not with respect to color changes of the fixation cross, the statistical

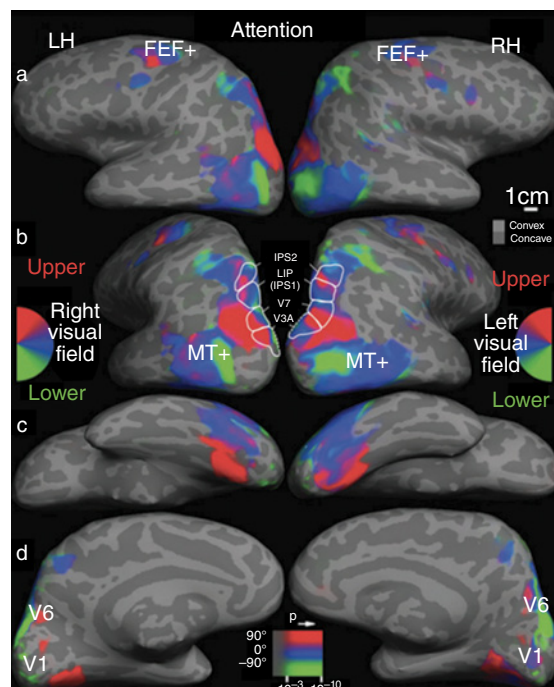


FIGURE 6.9.C Retinotopically sensitive cortex in the *Attention* condition.

map illustrated in *Figure 6.9.B* tells us nothing about object-based attention. When you've finished reading this chapter, come back to this page and think about how one might reanalyze these data to reveal something about object-based attention.

- *Attention to peripersonal vs. extrapersonal space.* This dimension was not engaged by this study. It will get a lot of attention in *Chapter 7*.
- *Competition.* This dimension was not engaged by this study, but will feature importantly in this chapter's section on *The Neurophysiology of Sensory Attention*.

Finally, when considering *Figure 6.9*, there is an important methodological/interpretational admonition to be made. (This is, after all a textbook, right? Please indulge the author in this lapse into heavy-handed didacticism) Think back to *Figure 6.5.B*. Just as with that figure, every part of the brains displayed in *Figure 6.9.B* and *Figure 6.9.C* are active (otherwise, they'd be dead!). "All" that *Figure 6.9.B* and *Figure 6.9.C* are showing us is the places where *some* of the activity fluctuates in a way that indicates an influence of the

experimental manipulation. Furthermore, all that they are showing us is the places where the researchers' analytic methods are able to detect this fluctuation. (I.e., through no fault of theirs, it's possible that additional regions/systems are also engaged by their manipulation, but that this engagement, for example, takes place at too fine a scale for fMRI to detect, or manifests via a mechanism – such as oscillatory coherence in a particular frequency band of the LFP – that fMRI simply doesn't measure.)

Now that I've gotten that off my chest, and we have seen how at least some of the clinical features of neglect syndrome can be understood in terms of the functional organization of the dorsal visual stream, let's consider, at a finer level of mechanistic detail, the kinds of computations that are carried out in the posterior parietal cortex.

Coordinate transformations to guide action with perception

Although *Figure 6.9* indicates that many areas of occipitoparietal cortex are organized retinotopically, they, too, follow the principle of hierarchical processing that we saw with the ventral visual processing stream in *Chapter 5*. Different from much the ventral stream, however, is the fact that many dorsal stream regions are characterized by multisensory integration, in particular, the integration of proprioceptive with visual information. Consider the following: as I sit writing this section, I occasionally lift my right hand from the keyboard to reach for the glass that is located to the right of the keyboard. Sometimes I turn my head to look at the glass before reaching for it, sometimes my face remains directed at the screen and I move my eyes to fixate the glass, and sometimes neither head nor eyes are moved, because while both are directed toward the screen, the glass occupies the outer edge of my right visual field. In each of these three scenarios, the glass is in the same location in allocentric coordinates (i.e., it's always in the same location on the arm of the couch in my living room), yet information that my nervous system processes about its location is different in all three. In the first two scenarios, it occupies the same retinotopic coordinates, but different head-centered coordinates, in that it is in a different location relative to my head

(straight ahead in scenario 1, at “2 o'clock” in scenario 2). In the second and third scenarios, it is at the same head-centered coordinates, but in different retinotopic coordinates. And in each of these three, what ultimately matters is that the glass is located in precisely the same location relative to my hand. What this example illustrates is that virtually all of the actions that we carry out over the course of our lives, certainly all of our sensory-guided actions, entail an ongoing process of translation between sensory-based and motor effector-based coordinate frames. To accomplish this, the nervous systems needs to know the position of each part of the body relative to the others (e.g., *Are the eyes pointing right or straight ahead?*). And recall from the final section of *Chapter 4* that this proprioceptive information is supplied by muscle stretch detectors.

As we shall review here, an important function carried out by the posterior parietal cortex is to compute the **coordinate transformations** that are necessary to know where an object is located relative to different parts of the body. The study illustrated in *Figure 6.9* provides no insight about how these coordinate transformations are computed, because it didn't involve movement of the body. Many studies that have provided important information about these processes have come from electrophysiological recordings of neurons in the awake, behaving monkey.

Extracellular electrophysiological studies of monkeys

Area 7a of the posterior parietal cortex (PPC), and the immediately superior intraparietal sulcus (IPS; see *Figure 5.5.A*) have been the focus of intensive investigation by neurophysiologists since the 1980s. In subsequent chapters, we will consider the role of these areas in such functions as the control of movement, the control of attention, and decision making. Presently, however, we'll focus on the role of these regions in the integration of visual and proprioceptive information. Neurophysiologist Richard Andersen and colleagues have studied this by first finding the retinotopically defined receptive field of a neuron in PPC, then systematically varying the location at which the monkey has to fixate, all the while keeping a visual stimulus in that neuron's receptive field (*Figure 6.10*). A property of these PPC neurons that

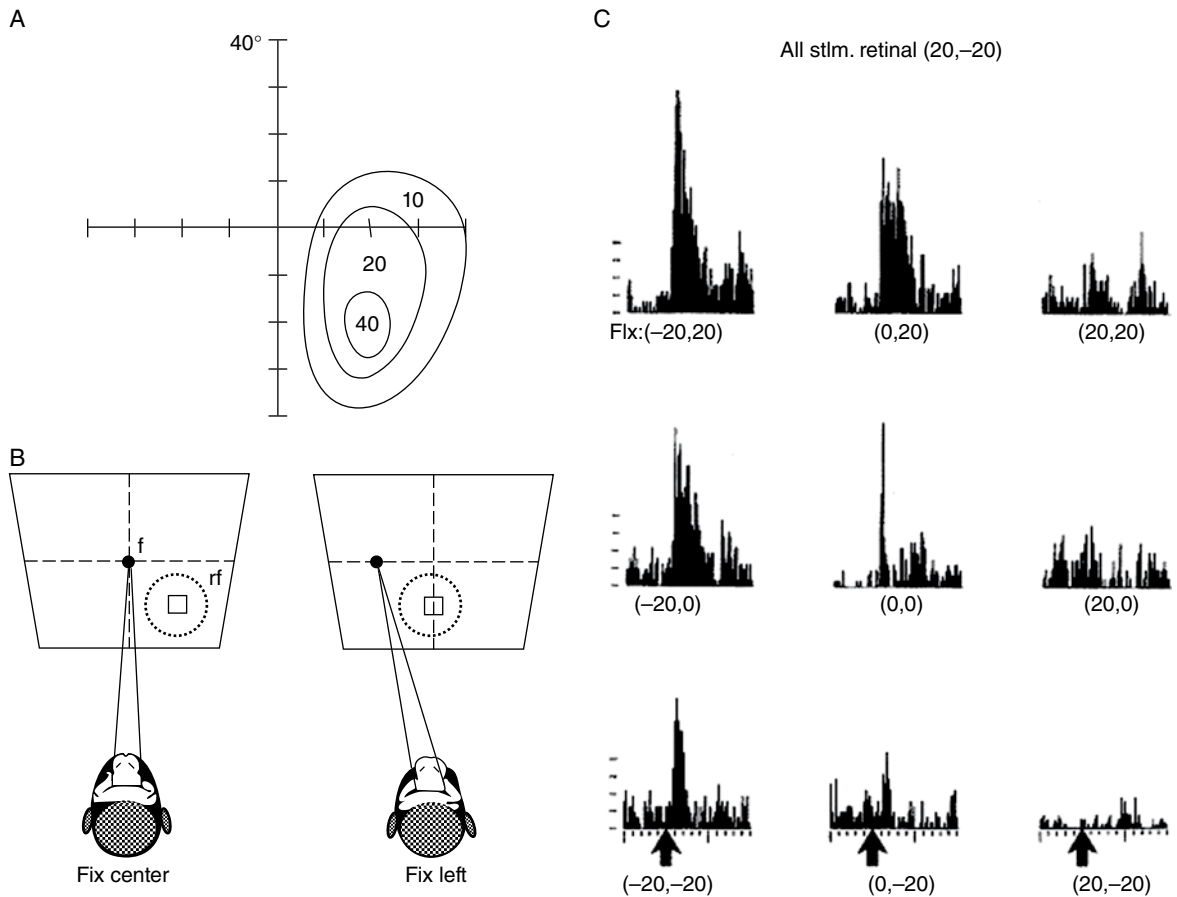


FIGURE 6.10 Eye position modulates spatial signaling of PPC neurons. **A** illustrates a contour plot of the receptive field of a neuron in area 7a, with the numbers indicating firing rate when a stimulus appears within each ring. **B** shows a monkey fixating two locations, with a stimulus presented within the receptive field for both fixations. **C** shows the responses of this neuron to the onset of the stimulus (arrow) in its receptive field when the animal is fixating at each of nine locations. Source: Andersen, Essick, and Siegel. 1985. Reproduced with permission of AAAS.

this procedure reveals is that the gain of their response is modulated by eye position. *Figure 6.10*, for example, illustrates a neuron with a retinotopic receptive field centered 20° to the right and 20° below fixation, and with the additional property that the magnitude of its response is sensitive to the position of the eyes within the head, with maximal gain when the animal is looking up and to the left. What this means is that, in effect, this neuron is conveying information not just about the location of a stimulus relative to the retina, but also about the location of a stimulus relative to the

head. Note that, in *panel B* of this figure, although the stimulus is in the same retinotopic location during “Fix center” and “Fix left” displays, it is in a different location relative to the animal’s head, which is in the same location for both conditions. As alluded to at the end of *Chapter 4*, coordinate transformations such as this are critically important, because an organism needs to know where a stimulus (say, a predator, or prey) is located relative to its body, regardless of the position of its eyes in their sockets, or of its head on its neck (*Tip 6.3*).

Tip 6.3

Andersen, Essick, and Siegel (1985) also point out that transformation of visual percepts into body-centered coordinates may be important for our sense that our visual world, and the objects in it, is stable, despite the fact that we move our eyes an average of three times per second.

Data such as those illustrated in *Figure 6.10* indicate that neurons in area 7a of PPC must receive information about eye position from somatosensory cortex, in addition to retinotopic visual information. (Eye position is computed from stretch receptors that are sensitive to the state of contractedness vs. relaxedness of the muscles that control the movement of the eyes.) Evidence for such integration of information from different sensory domains, and the coordinate transformation that this allows, suggest that the PPC is well positioned to mediate the transition between sensory attention and motor intention. Indeed, more explicit evidence for PPC involvement in motor control comes from delayed-response tasks. In these tasks, a cue stimulus is briefly presented to identify for the subject (in the present example, a monkey, but the same procedure is also used with humans) the region on a screen to which it will need to make an eye movement (a **saccade**) or a reaching movement. The subject has been trained, however, to withhold this movement until the offset of the fixation point, which lags behind the flash of the cue by several hundred milliseconds, or longer. *Figure 6.11* illustrates that extracellular recordings collected during a task that varied whether the animal was to make an eye movement or a reach, identified two regions of the PPC that showed preferentially elevated delay-period activity as a function of which body part was to be used on that trial. (Note that, because the cued location was the same on both types of trial, these neurons were presumably computing information relevant to their preferred motor effector, rather than to a more abstract representation of where in space the cue had been presented.) In *Chapter 7* we will return to some implications of findings such as these for consideration of whether the dorsal stream is better construed as

processing *where* a stimulus is located, or as processing *how* the organism is going to interact with it.

The coupling of perception with motor control in the PPC becomes even more concrete when one considers the effects of electrical microstimulation.

Microstimulation of PPC produces eye movements (and more)

The multiple panels of *Figure 6.12* present some of the results from a heroic study by Thier and Andersen (1998) of the effects of microstimulation in several regions of the PPC. Recall from *Chapter 1* that an early application of this approach was used in the 1880s to establish (a) that motor control is localized to the frontal cortex and (b) on a more fine-grained level, that there's a topography to motor cortex, such that different regions of frontal cortex control different parts of the body. The intent here is not to go into detail about the control of eye movements, which will be a focus of *Chapter 8*, but to further emphasize the point that the distinction between perception, attention, and action loses its clarity in the PPC (*Tip 6.4*). Microstimulation of the PPC produced two principal classes of microstimulation-evoked saccade. The first, "*modified vector saccades*," are named in reference to "*vector saccades*" which are of a fixed amplitude and direction, regardless of starting eye position (we'll encounter these again in the FEF in *Chapter 8*). Those illustrated in *Figure 6.12.B* are "*modified vector saccades*" in that whereas their direction is invariant, their amplitude depends on the starting position of the eye. The other principal class of saccades observed was "*goal-directed*" saccades, in which the same endpoint in x-y space is achieved, regardless of starting point (*Figure 6.12.C*). Many of the locations producing goal-directed saccades also produced what the authors summarized as "*noneye movements*" – microstimulation-evoked movements of the face, pinnae (a.k.a., ears), shoulders, etc. They

Tip 6.4

If you're in a facetious mood, you might say that this region of the brain does not honor the tidy categories that scientists have devised to explain the functions of the brain

FIGURE 6.11 The response of PPC neurons to delayed response with the eyes vs. the hand. Source: Snyder, Batista, and Andersen, 1997. Reproduced with permission of Nature Publishing Group.

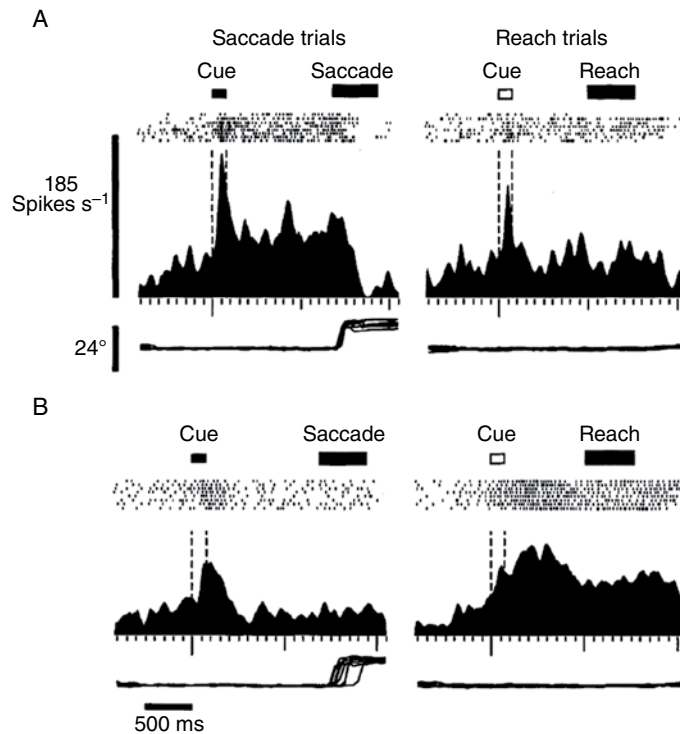


FIGURE 6.11.A Row **A** illustrates the responses of a neuron to a red-colored cue (filled bar in this diagram), signaling that a saccade will be required, and to a green-colored cue (open bar), signaling that a reach will be required. The 8 rows of rasters (tick marks) each illustrate the timing of action potentials recorded during a single trial; the histogram below the average from those trials; and the traces below that the position of the eye along the vertical meridian as a function of time. Note the near-instantaneous “jumps” in eye position characteristic of a saccade, as opposed to the smooth movement of tracking a moving object. This neuron responds more during the pre-saccade than the pre-reach delay period. Row **B** illustrates responses of a “reach-specific” neuron on the same trials.

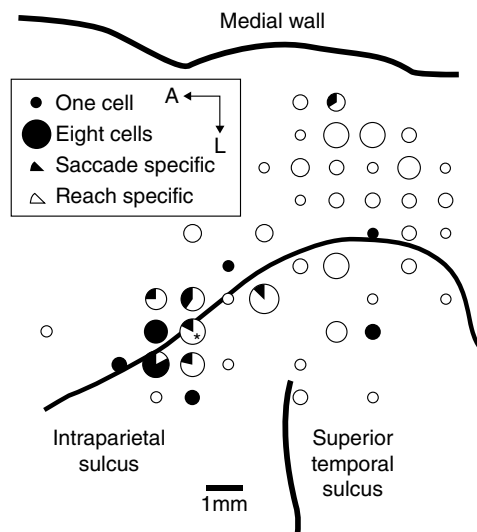


FIGURE 6.11.B The anatomical segregation of saccade-specific neurons (filled circles, found primarily along the walls of the intraparietal sulcus) and reach-specific neurons (unfilled circles, found more posteriorly and medially). “A” = anterior; “L” = lateral; “medial wall” = midline

FIGURE 6.12 Microstimulation of PPC evokes eye movements. Source: Thier and Andersen, 1998. Reproduced with permission of The American Physiological Society (APS).

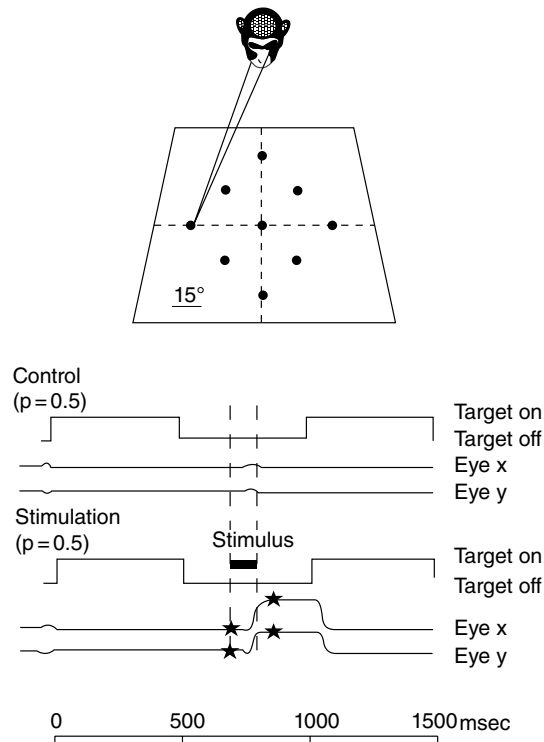


FIGURE 6.12.A Top image shows a monkey fixating one of nine target locations, only one of which was illuminated to begin a trial. On “control” trials, the light was on for 500 ms, then extinguished for 500 ms, then reilluminated. Eye position traces (illustrating position of one eye in x and y dimensions) indicate that the animal maintained fixation throughout trial. On “stimulation” trials, microstimulation was delivered during period between the two dotted lines. Saccade metrics were measured during period marked by two stars. Note that, after reillumination of the target, the eyes returned to starting position.

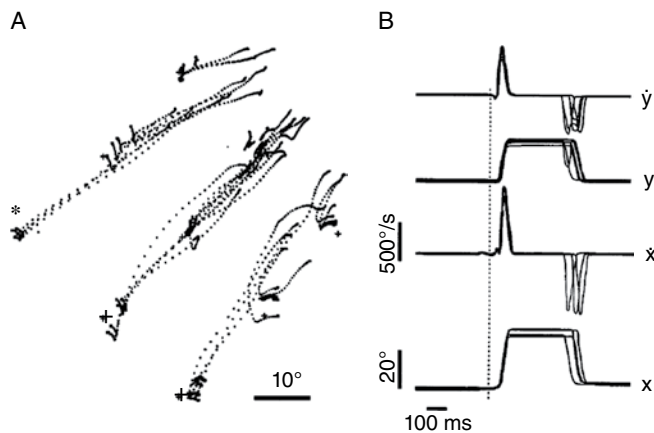


FIGURE 6.12.B Panel **A** illustrates “vector saccades” produced from each of the nine starting points by stimulating in area LIP. It is characteristic (but not universal) that the amplitude of the evoked saccade is larger for starting position further from the direction of the saccades. Panel **B** illustrates the kinematics for the saccades beginning from the target in A indicated by the asterisk. The traces show, from top to bottom, velocity of the eye in the y direction as a function of time, location of gaze in the y dimension as a function of time, and the analogous measures in the x dimension.

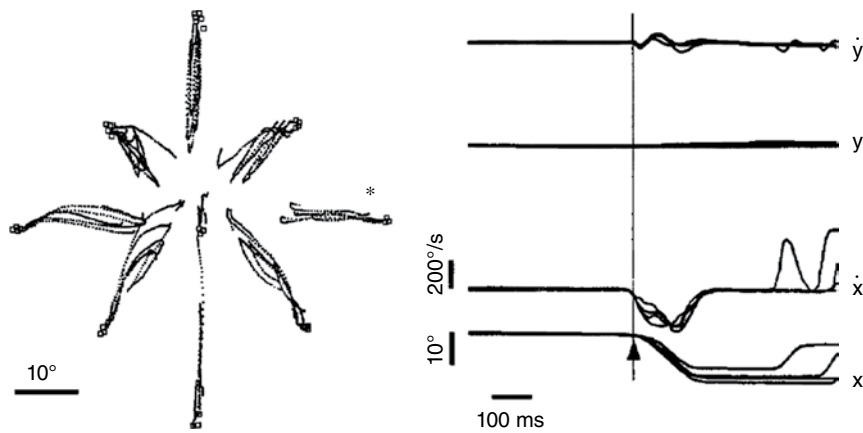


FIGURE 6.12.C “Goal-directed” saccades produced by stimulating white matter below the fundus of the IPS. All conventions same as *Figure 6.12.B*.

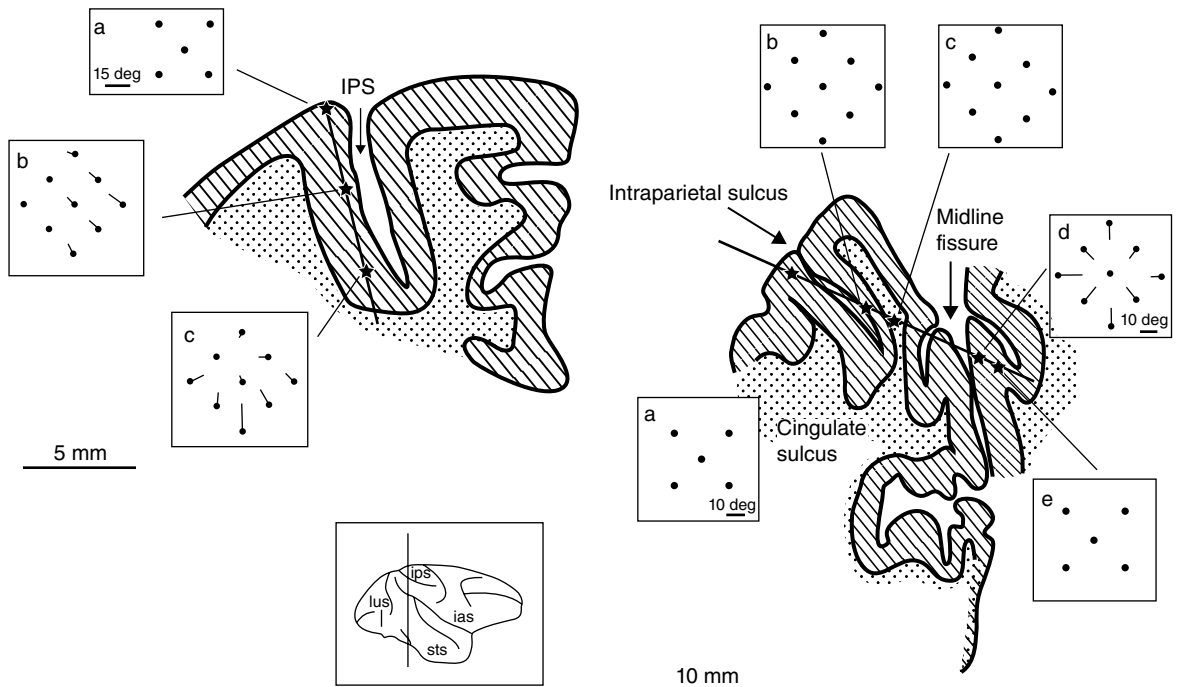


FIGURE 6.12.D Illustration of findings from two electrode penetrations, with the path of the electrodes illustrated by the black lines. Each star indicates a microstimulation site, and the boxed “call out” diagrams illustrate the vector-averaged saccades produced by microstimulation in relation to each target.

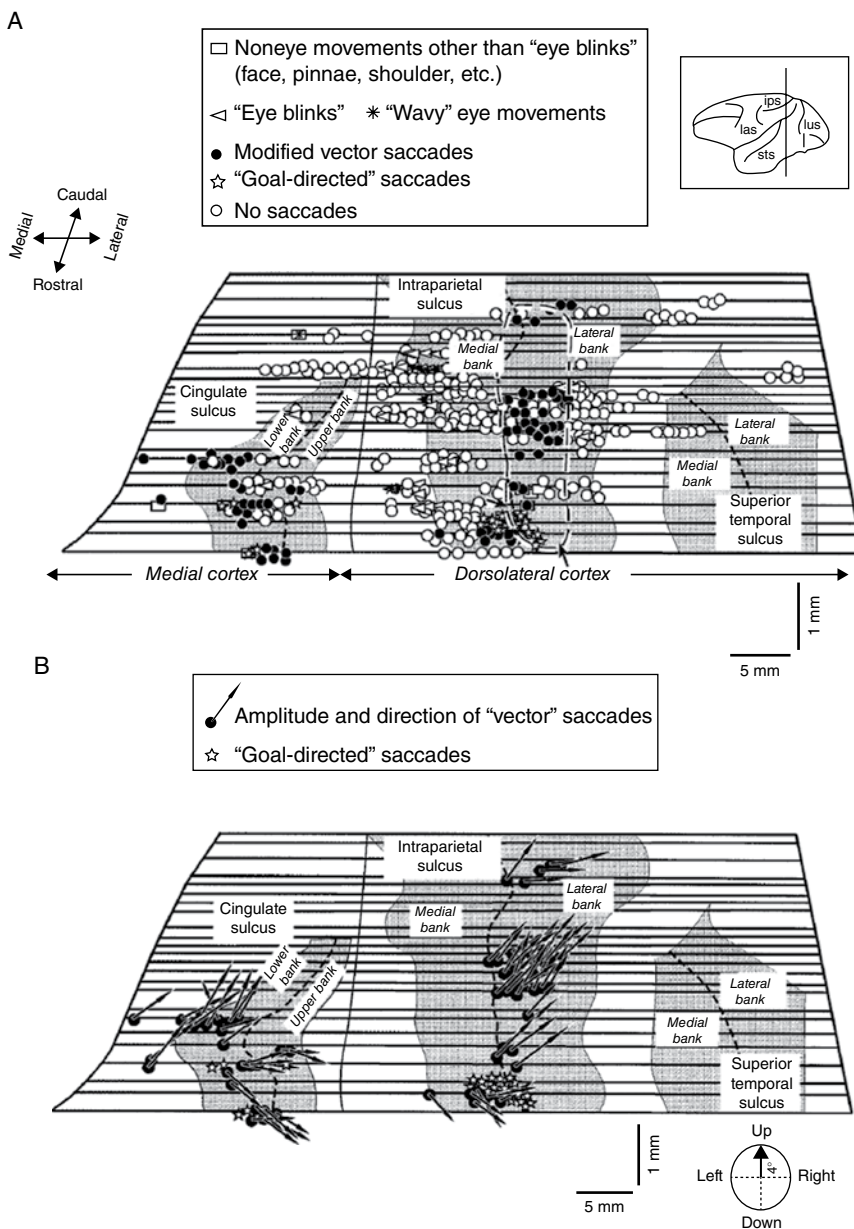


FIGURE 6.12.E Two flat maps summarizing regional organization of microstimulation effects. Map **A** classifies the results of microstimulation at each location where it was applied (see text for definitions). Map **B** illustrates the amplitude and direction of “modified vector saccades” produced at each site identified in Map **A**. The convention for both maps is that the walls of sulci are shaded, a dashed line indicating the fundus. The solid line dividing “medial cortex” from “dorsolateral cortex” is the point of inflection (the “crease mark”) that defines the midline. (The elongated, longer-dashed enclosure relates to a second monkey, and is not relevant to our purposes.)

speculated that these may indicate involvement in the coordinated movement of multiple parts of the body (e.g., turning the head while moving the eyes while reaching, all toward the same “goal”). Note that the segregation of regions within the IPS producing these two classes of movements suggests differential anatomical connectivity with other cortical and subcortical regions.

Although these two types of saccade are qualitatively different from one another, both are indicative of reference frames that incorporate information in addition to retinotopy. In this section we’ve seen how coordinate transformations are vital to an organism being able to control its actions in appropriate ways. Next, we’ll turn to what might be “the mother of all coordinate transformations,” which is from the myriad body-centered representations that are necessary for moment-to-moment behavior, to the “body-free,” world-centered (or “allocentric”) representations that are a central (and perhaps necessary) component of our stored knowledge about the world.

FROM PARIETAL SPACE TO MEDIAL-TEMPORAL PLACE

Let’s return for a moment to the patients of Bisiach and Luzzatti (1978; *Figure 6.4*). The fact that they were able to correctly conjure up several landmarks surrounding all sides of the piazza, despite the fact that both required two perspectives from which to do it, means that their long-term memories contained a complete representation of this locale, one in which the position of each element relative to the others was stable – an allocentric representation. Their deficit evidently came from the fact that, in order to think about this representation, they had to do so from an egocentric perspective. First, let’s consider a neural substrate for this allocentric representation of space. In the final section of this chapter, we’ll take up how one might begin to address the notion of “thinking about” it.

Place cells in the hippocampus

Important insights about the formation of allocentric representations of space have come from research in the rat, much of it propelled by the groundbreaking research of the US-born, Canadian-trained,

UK-based neurophysiologist John O’Keefe and colleagues. The initial discovery was made while acquiring extracellular recordings from the dorsal hippocampus of a rat while it freely moved about, exploring an elevated maze (essentially a small tabletop with “runway” arms, and no side walls; *Figure 6.13*). What they found was that whereas some pyramidal neurons of the hippocampus responded to certain behaviors, such as grooming or sniffing and whisking, others showed a qualitatively different profile, seeming to encode where the animal was in the environment, independent of the orientation of its body. The difference was so striking that they categorized hippocampal pyramidal neurons into two groups: “the *displace* units are those whose firing pattern relates to the behaviour of the animal, irrespective of where it occurs in an environment, while *place* units are those whose firing pattern is dependent on the location of the animal in an environment” (O’Keefe and Nadel, 1978, pp. 196–197; italics added). Their choice to label *displace* units in reference to the property that they did not display indicates which of the two they found more interesting.

As shown in *Figure 6.13*, “place cells,” as they have come to be known, are qualitatively different from cells with body-centered receptive fields, because the fundamental piece of information that they represent is where the animal is located in space, not what it is currently seeing or feeling (*Tip 6.5*). Thus, the neuron illustrated in *Figure 6.13* has a *place field* (by analogy to receptive field), in that it fires selectively whenever the animal passes through or stops in this location, regardless of the direction in which it is facing. It does not fire, for example, when the animal leaves B and turns right to enter C, even though this movement sequence is almost identical to that required by “Run A→B.” Another interesting property is that the magnitude of this neuron’s discharge, in a pattern reminiscent of the gain fields of the lateral intraparietal area (LIP), is influenced by factors other than place in and of itself. For example, the response is visibly stronger when the animal is entering the B arm than when leaving, and seems to be stronger when passing through the place field than when sitting in it.

Note that, in isolation, a single place cell’s output wouldn’t be very informative to the organism, in that

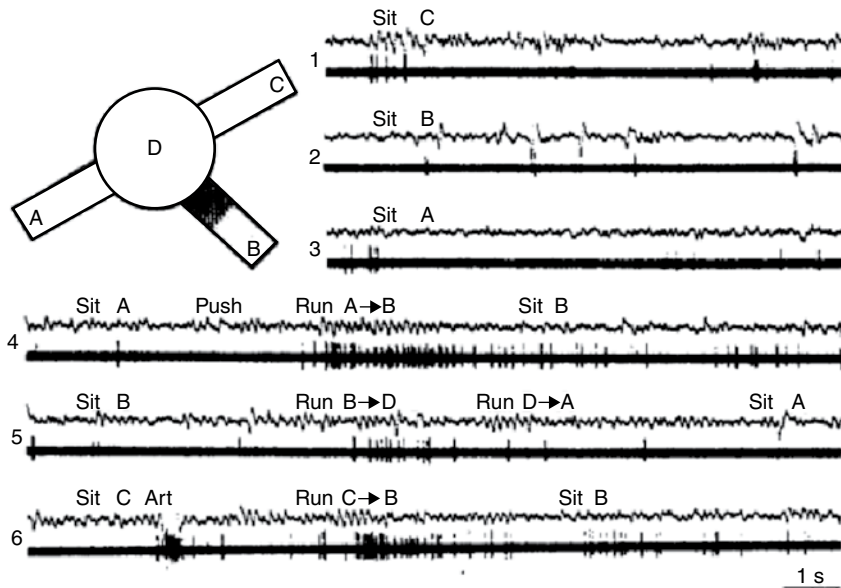


FIGURE 6.13 A place cell recorded from the dorsal hippocampus of the rat. The image on the upper left is a schematic top-down view of a three-arm maze, with the shaded area in B the location of the cell's place field. A pair of electrode traces is shown from each of six epochs of behavior, the upper the LFP (band-pass filtered to only display the theta band), the lower the spiking. Note the neuron produced place-related bursts of activity regardless of whether the animal had turned right to enter the B arm (epoch 4), turned left to enter the B arm (epoch 6), sat in the B arm (epochs 4 and 6) or was leaving the B arm (epoch 5). (The label "Art" in epoch 6 identifies an artifact in the recording.) Source: O'Keefe and Nadel, 1978. Reproduced with permission of Oxford University Press.

Tip 6.5

Subsequent research has shown tantalizing results suggesting that these neurons might also be active when the animal is only "thinking" about that location in space, as might occur during dreaming!

it can only signal, for example, that *I am currently in the corner of my living room near the fireplace*, or *I am not currently in the corner of my living room near the fireplace*. Thus, when walking around in my kitchen, say, or from my bedroom to that of my son, this neuron, through its silence, would be continually, implicitly, conveying that *I am not currently in the corner of my living room near the fireplace*. Thanks a lot. But, of course, individual neurons don't operate in isolation. Indeed, as Figure 6.14 illustrates, a population of place cells, by

each representing a different location, is ideally suited to represent an environment. Indeed, one can see how, just as is the case with receptive fields in the visual or somatosensory system, the smaller the hippocampal place field, the finer-grained the representation of that region of space. (*Web Link 6.4* builds on these ideas to address "How your brain tells you where you are.")

How does place come to be represented in the hippocampus?

The detailed allocentric representation of a locale with which one is highly familiar (say, the Piazza del Duomo for natives of Milan) comes into existence over time, stitched together from multiple egocentric "snapshots." Only by having stood on the steps of the cathedral and looked out across the plaza, having strolled along the shops, having had a coffee at one of the outdoor cafes, and so on, could one construct a

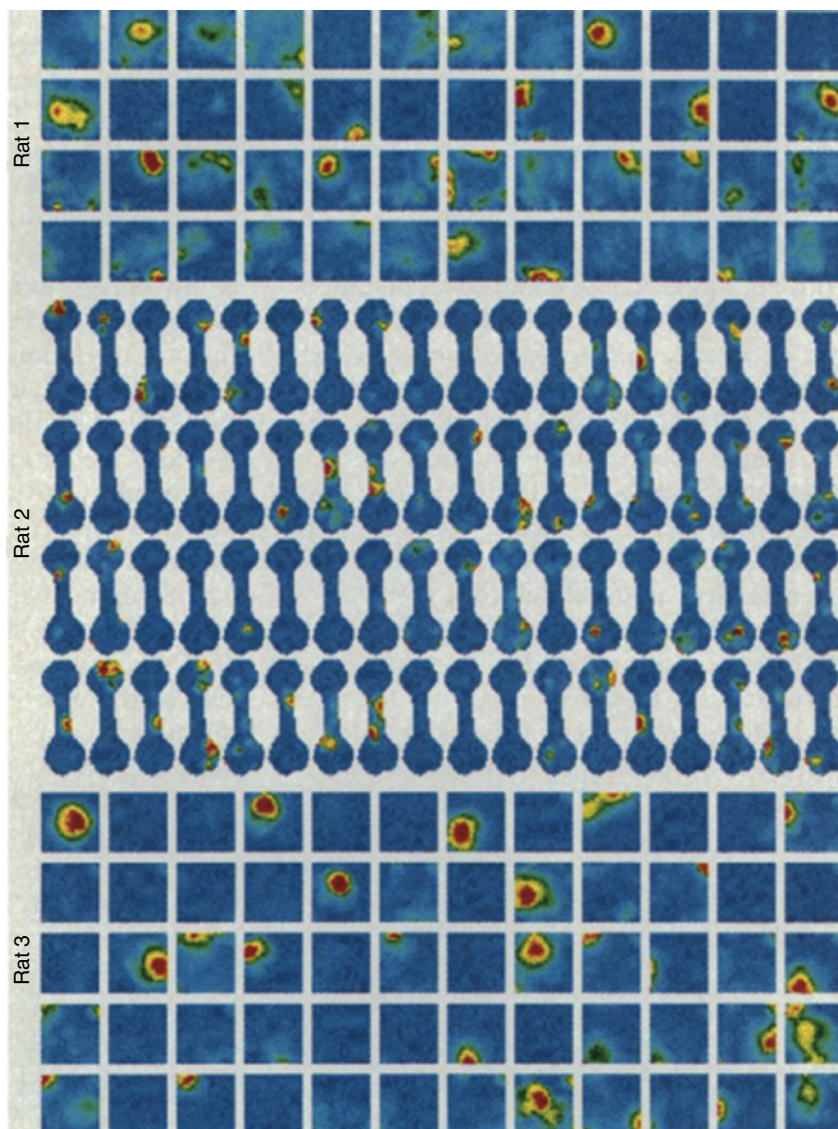


FIGURE 6.14 Place fields of multiple hippocampal pyramidal neurons from three rats. Each animal's panel shows multiple top-down views of the environment that the animal was exploring (a square-shaped enclosure for rats 1 and 3, a dumbbell-shaped enclosure for rat 2). Each individual square or dumbbell image represents the spatial tuning of a single neuron, the heat map indicating the neuron's firing rate at each location of the enclosure (range: blue=0 Hz; red= ≥ 10 Hz). These illustrate how, by sampling across a population of place cells, the hippocampus can "map" the entirety of a locale. *Web Clip 6.3* includes a movie of an experiment recording from place cells. Source: Wilson and McNaughton, 1994. Reproduced with permission of AAAS.

representation of how each of the discrete elements in this locale relate to one another. Because these egocentrically defined "snapshots" each depend on processing in the dorsal stream, it is reasonable to ask

how it comes to pass that allocentric space is represented in a structure that lies along the rostromedial wall of the *temporal* lobe (*Figure 2.2*). From *Figure 5.5.A*, one might conjecture that the hippocampus is at

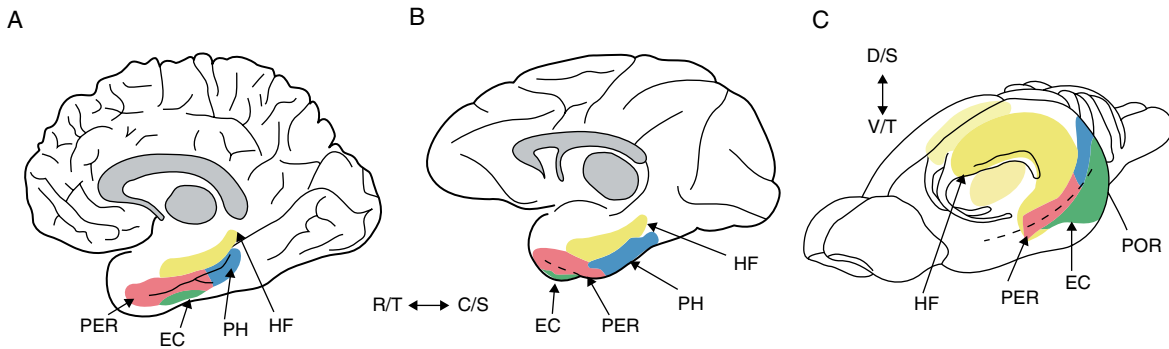


FIGURE 6.15 Location of the hippocampal formation (HF) and adjacent cortex in the human (**A**), the rhesus macaque (**B**), and the rat (**C**). The cortical areas are the perirhinal cortex (PER), the entorhinal cortex (EC), the parahippocampal cortex (PH) and, in the rat, the postsubiculum (POR), a homologue to primate PH. The convention for directional labels is general anatomical label/hippocampus-specific label, thus: R/T=rostral/temporal; C/S=caudal/septal; D/S=dorsal/septal; V/T=ventral/temporal. Note that area 36, which *Figure 5.5.B* shows to be the temporal-lobe target of PPC projections, corresponds to the blue-colored POR/PH in this figure. An intriguing disorder of ego- and allocentric processing that results from damage to this region is described in the *Further Reading* by Aguirre (2003). Source: Burwell and Agster, 2008. Reproduced with permission of Elsevier.

the terminus of the ventral stream, located adjacent, as it is, to vAIT. Indeed, this reasoning wouldn't be incorrect; just incomplete.

Figure 5.5.B indicates that, in fact, BA 7 in the posterior parietal cortex and AIT are equidistant from the hippocampus in terms of connectivity, each only two synapses removed. To understand how this organization came to pass, it's useful to return to the rodent. In both the rodent and the primate, the hippocampus is slightly curved, like a banana. Whereas in the primate its long axis parallels that of the temporal lobe, in the rodent it is oriented dorsoventrally, effectively “cupping” the caudal surface of the thalamus. Thus, the dorsal hippocampus (where O'Keefe and colleagues initially discovered place cells) lies directly below the cortical surface of the parietal lobe (*Figure 6.15*). Scanning across species in *Figure 6.15* makes it seem plausible that, over the course of evolution, the primate hippocampus has “migrated” ventrorostrally from a previous dorsoventral orientation (*Tip 6.6*), and “taken its parietal connections along with it.”

From place to long-term memory?

In *Section III* of this book we'll see that the hippocampus is also implicated in the formation and subsequent processing of long-term memory (LTM). We get a

Tip 6.6

Around the time that I am writing this section, paleontologists have pieced together fossil evidence of the oldest known common ancestor of all modern mammals, a decidedly rodent-looking four-legged critter (see *Further Reading* reference to O'Leary et al., 2013).

preview of this in *Web Clip 6.3*, which highlights a question being pursued by many cognitive neuroscientists: *Are the hippocampus's roles in spatial cognition and in LTM related in some fundamental way, or are they independent functions that just happen to be supported by the same brain structure?*

THE NEUROPHYSIOLOGY OF SENSORY ATTENTION

The deficit [of the patients of Bisiach and Luzzatti (1978)] evidently came from the fact that, in order to think about this representation, they had to do so from an egocentric perspective. It follows from this sentence from the previous section that cognition often requires us to select

just a subset of the totality of information that's available to us, in order to submit this subset of information to the processing that will enable us to achieve a behavioral goal. The operations of selection and the prioritization for subsequent processing are operations ascribed to the function of *attention*. Among the influential late-nineteenth-century psychologists whose research and writing laid important foundations for contemporary cognitive neuroscience (see the concluding section of *Chapter 1*) was William James (1842–1910), who famously wrote that “Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought” (James, 1950, pp. 403–404). This nicely captures our intuition. The difficulty of understanding the cognitive and neural bases of attention, however, was cheekily captured by cognitive scientist Harold Pashler, who wrote roughly a century later that “no one knows what attention is, and ... there may even not be an ‘it’ to be known about (although of course there might be)” (Pashler, 1999, p. 1). Here, we'll scratch the surface of this intriguing, notoriously slippery phenomenon that will be relevant for almost every aspect of cognition and behavior that we consider from here on out.

A day at the circus

The metaphor of attention as a spotlight has been a useful one for researchers, in that it captures the phenomenon of directing a resource (light) so as to prioritize one out of many possible targets. At the circus, for example, the lighting director may want the audience to attend to the diversionary pratfalls of a juggling clown, rather than to the trapeze artist who is climbing a ladder up to a high platform, but isn't yet performing, or to the animal handler wrangling the recalcitrant tiger back into its cage, after it has jumped through the Hoop of Fire. To accomplish this, the spotlight is shined on the clown, thereby making him brighter and easier to see than the climbing acrobat or the retreating tiger. What if Clown 1 is joined by two or three additional jugglers? The area being illuminated by the spotlight can be enlarged or reduced, depending on the number of clowns performing. When it's time for the

trapeze performance, the spotlight can jump from the floor, where the clowns were juggling, to a platform high overhead – the platform is currently empty, but the director wants the audience to be attending to the location onto which the trapeze artist will soon make her dramatic, graceful entrance.

This scenario captures two properties of attention that we will briefly consider here: the function of *increasing the salience* of one or more selected stimuli; and the distinction between *attending to an object* versus *attending to a location*. A third, the ability to change the aperture of attention, will be considered in *Chapter 13* (under the rubric of *capacity limitations* of cognitive functions). A final, critical element in our circus scenario is the person who controls the spotlight, turning it on and off, and pointing it first at the clowns, then at the platform. The invocation of the spotlight operator highlights the distinction between the *site* of attention's effects (in our scenario, the clowns, then the platform), versus the *source* of attention (the spotlight itself, which generates the bright light, and the person controlling its operation). Our present focus will be on the site of attention, its source(s) to be considered in *Chapter 8*.

Attending to locations vs. attending to objects

One place where we've already seen the effects of attention was in *Chapter 4*, when considering the difference in the AEP when the subject attended to a series of beeps, versus when he read a book and ignored the beeps. In that example, because the effect of attention was to increase the magnitude of the N1 and P2 components of the AEP (*Figure 4.6*), we can say that the site of attention's effects included primary auditory cortex. And what might have caused this “amplification” of these ERP components? There are at least two plausible explanations. One is that the neurons responding to the afferent auditory information may have done so more vigorously in the count-the-beeps condition than in the book-reading condition. A second is that more neurons may have been recruited when the subject attended to the auditory channel (*Tip 6.7*). And recall from *Chapter 4* that loudness is encoded in the auditory pathway by firing rate – the louder a sound, the higher the rate of firing

Tip 6.7

Note that there are also two possible ways by which *recruiting more neurons* might be accomplished, and not all scientists will agree on which account is better. The first is that more neurons in A1 begin firing in response to the click. The second is that the phase of the ongoing membrane oscillations of more A1 neurons is reset by the click-evoked volley of action potentials arriving from the MGN

of neurons representing that sound's "places" along the basilar membrane.

Thus, by either of these two accounts, a reasonable supposition is that paying attention to a particular sensory channel may be equivalent, in neural terms, to increasing the energy of the stimulus itself (such as by making it louder, or brighter). The phenomenon of an attention-related boost in the magnitude of sensory processing signals is also seen with the visual-evoked potential (VEP) and somatosensory-evoked potential (SEP). *Figure 6.16* illustrates a test of visual

spatial attention, in that subjects fixate centrally but focus their attention on a location and wait for events to occur there. (This is referred to as **covert attention**. **Overt attention** would be to look directly at the

FIGURE 6.16 EEG of spatial-based visual attention. Source: Di Russo, Martínez, and Hillyard, 2003. Reproduced with permission of Oxford University Press.

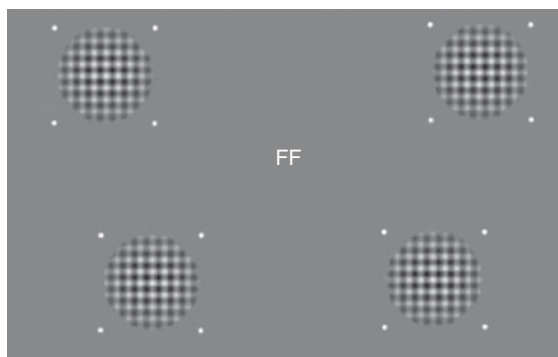


FIGURE 6.16.A The display: subjects were instructed to attend centrally, and count the number of flashed checkerboard stimuli appearing in just one of the four locations demarcated by white dots. (This is a variant of the famous Posner (1980) spatial cuing paradigm.)

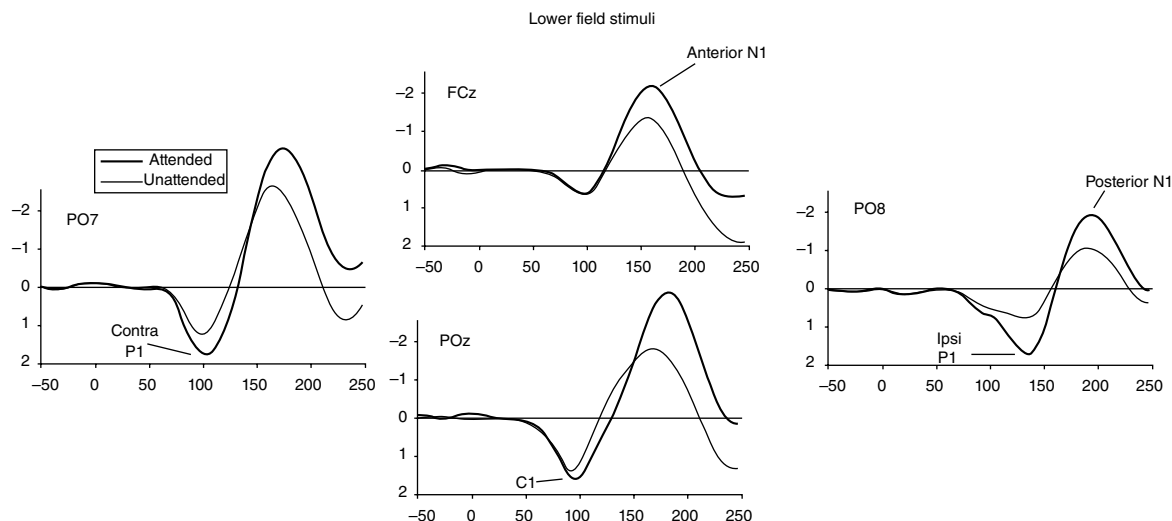


FIGURE 6.16.B ERPs associated with stimuli presented to the lower right location, when that location was *Attended* and when it was *Unattended* (i.e., when visual presentation was identical, but subjects were attending to the left visual field location). The four plots show the signal recorded at electrodes FCz, PO8, POz, and PO7 (in relation to *Figure 3.9*, the FCz is located midway between Fz and Cz, and the PO electrodes midway between the P and O electrodes).

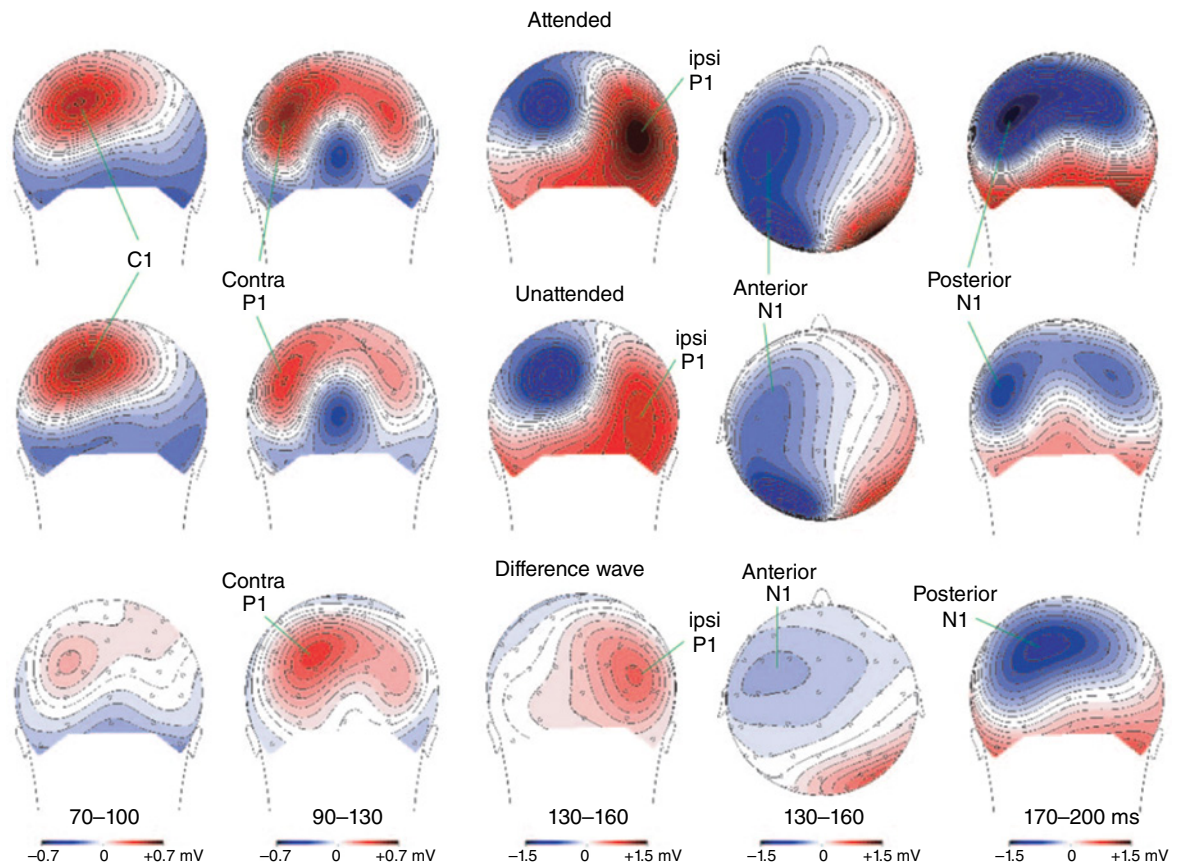


FIGURE 6.16.C Four “snapshots in time” of the ERP from Figure 6.16.B as it distributed across the scalp (as viewed from behind, except fourth column from left shows from above). The top and middle rows show the ERP vs. baseline, the bottom shows the subtraction images of [Attended – Unattended]. Although recorded from an array of 64 discrete electrodes, interpolation was applied to generate this continuous 2-D voltage map.

location of interest.) From it, one can see that the effect of spatial attention on early components of the VEP is to magnify them. Further, it seems to do so in a particular way. Looking at the traces from the PO electrodes, it's as though the voltage of the *Attended* ERP at each time point was computed from the corresponding time point of the *Unattended* waveform by multiplying the latter by some constant value (perhaps 1.7 at PO7, and a value closer to 2.0 at PO8). Based on this observation, in isolation, one might hazard that attention works by applying a multiplicative gain factor to sensory information processing. We'll return to this idea in a moment, after first

considering two additional pieces of evidence that complicate things a bit.

Whereas it's clear that we can attend to a location in space and “wait for something to happen” there, our intuitions, confirmed by decades of experimental psychology, tell us that we can also attend to an object, regardless of where it's located in space. (I do this, for example, when proudly admiring my daughter at a gymnastics competition, despite the fact that she's surrounded by a milling gaggle of many, many other young girls, all wearing similar leotards and hairstyles.) A clever ERP experiment designed by the same group featured in the previous paragraph, led by

electrophysiologist Steven Hillyard, directly compared the effects of spatial attention vs. those of nonspatial feature-based attention. (For our purposes, it's simpler to start by considering single features, rather than complex objects.) In it, subjects attended to a nonspatial feature of the stimulus, its color or its direction of motion, which, from trial-to-trial, appeared in one of the two visual hemifields (Figure 6.17.A). Comparison of the VEP for the attended vs. the unattended hemifield revealed the amplification of the P1, N1, and subsequent visual-evoked components that we have already seen (compare Figure 6.17.B to Figure 6.16.B). The comparison of the attended nonspatial feature vs. the unattended nonspatial feature, in contrast, revealed an ERP component that was not “merely” an amplification of the VEP (Figure 6.17.B). Instead, this “selection negativity” (SN), must correspond to the enhanced processing of the critical stimulus feature (i.e., of its redness, or of its horizontal-ness), rather than to the indiscriminant amplification of all activity in the visual system that seems to result from spatial attention. That is, it seems to reflect a neural process corresponding to the *selection* of one feature from the two-feature stimulus. It is important to note that although the shape and timing of the SN for attention to a color or to a direction of motion appear to be very similar, the two were extracted from electrodes on different parts of the scalp: the color SN was greatest at contralateral occipitotemporal sites, whereas the motion SN was greatest at bilateral temporal and parietal sites (Tip 6.8).

Tip 6.8

Since the 1950s different theories and different sets of behavioral data have been used to debate the question of whether attentional selection occurs *early* in sensory processing (i.e., at the level of feature detection, “before” meaning has been assigned the incoming sensory information) or *late* (after object recognition has occurred). Although word-limit constraints preclude a detailed accounting, the ERP studies summarized here are paradigmatic of how neuroscience measures can be used to address cognitive questions.

Mechanisms of spatial attention

Contradictory findings from fMRI vs. EEG?

Having established the distinction between location- vs. feature-based attention, let's now consider how they might work. Recall that Figure 6.16.B suggested a mechanism applying “multiplicative gain.” *Huh?* Consider, for a moment, your iPod (or whatever device you kids use to listen to your “music” these days): you set the volume to a larger number and the music is louder; you set it to a smaller number and it's softer. But what about during a silent gap in the music? Whether the volume is set to 2 or to 11, the silence will sound the same. This is because the products of $[2 \times 0]$ and of $[11 \times 0]$ are both 0. Your volume controller only influences the gain of the signal being produced by your music player, by *multiplying* it by some factor. It doesn't

FIGURE 6.17 EEG of feature-based visual attention. Source: Anllo-Vento and Hillyard, 1996. Reproduced with permission of Springer.

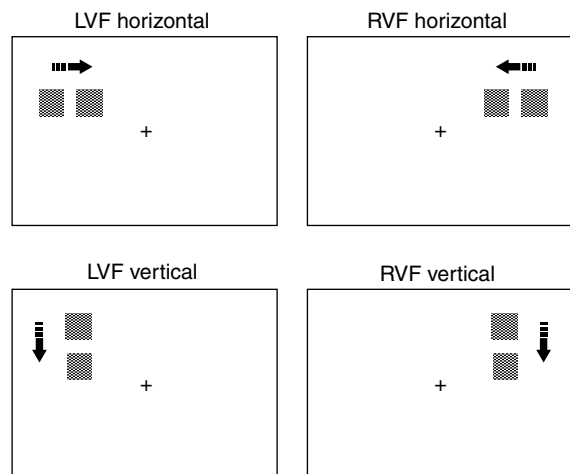


FIGURE 6.17.A There were three relevant task dimensions, color (red, blue) *direction* of motion (horizontal, vertical), and visual field (right, left). Each of the six possible stimulus types (e.g., red horizontal motion in the LVF) could occur on each trial, and two varied randomly on each trial. On any block of trials, subjects were to indicate when the *speed* of the moving stimulus was slower than typical, but only when this slowing occurred for the instructed stimulus type (red, blue, horizontal, or vertical) in the instructed visual field.

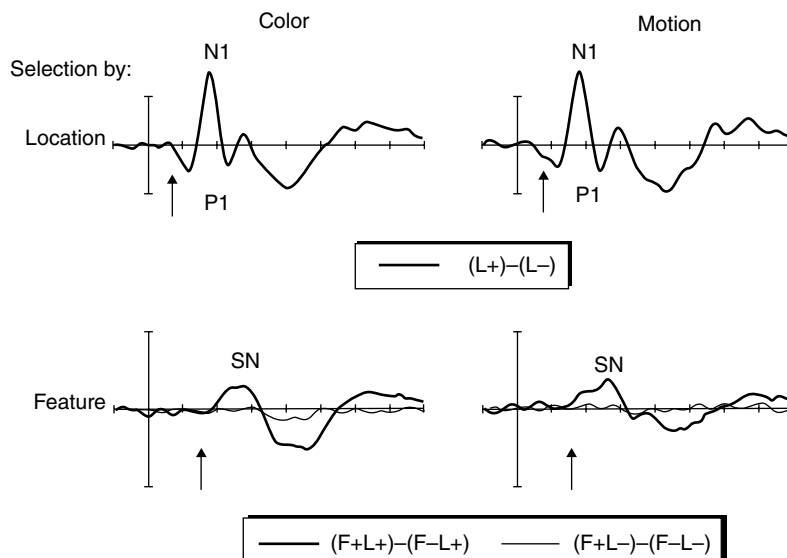


FIGURE 6.17.B ERP difference waves showing the effects of spatial attention (top row) and feature-based attention (bottom row). The waveforms in the top row are conceptually what would result if, in *Figure 6.16.B*, one were to subtract the “Unattended” (here, labeled “L-”) from the “Attended” (L+) waveforms at any electrode – they illustrate the extent to which the VEP is magnified by spatial attention while the subject was attending to a color (left-hand column) or a direction of motion. In the bottom row, the waveforms drawn in heavy lines illustrate the subtraction of trials when the instructed color (or direction) was present in the attended visual field (“F+L+”) versus when the uninstructed feature was present in the attended visual field (“F-L+”), thereby emphasizing the “selection negativity” (SN) component. The faint lines are the analogous subtractions from the unattended visual field. Arrows indicate the points at which attentional effects first achieved statistical significance.

add anything to the signal. (Now look back at *Figure 6.16.B* and note how the Attended and Unattended waveforms are identical until afferent visual signals get to V1, at which point the attended vs. unattended VEPs diverge in a multiplicative manner.) An *additive* process, on the other hand, might look like what is illustrated in *Figure 6.18*. This illustrates an fMRI study of both location- and feature-based attention, the former being our interest at the moment. Note how, in *Figure 6.18.C*, there is an increase in fMRI signal, relative to baseline, during the “expectation” period of the task. That is, although the “BL” and “EXP” periods of the task are identical with respect to visual presentation (empty gray screen) and overt behavior (fixation), the signal in brain areas representing the upper visual field nonetheless increases when spatial attention is directed there. This phenomenon is sometimes referred to as a “baseline shift,” in that “nothing has

happened yet” in the EXP portion of the trial, and so an analysis that ignored the cognitive state of the subject (“not attending” vs. “attending”) would conclude that the EXP period was a BL period with a level of activity that was shifted upward.

And so the comparison of the EEG studies of *Figures 6.16* and *6.17* vs. the fMRI study of *Figure 6.18* seem to suggest that spatial attention operates via a multiplicative or an additive mechanism, respectively. Is one wrong? Of course, it’s not that straightforward. We need to take into account that the two sets of experiments were intended to address different questions (“early vs. late selection” being an important question motivating the ERP studies (see *Tip 6.7*), “biased competition” (which we’ll take up presently) motivating the fMRI study), and that these two techniques are sensitive to different neurophysiological signals. Additionally, how the data are analyzed can determine what one is capable of detecting in one’s

data. Thus, although the ERP studies summarized here don't provide evidence for a baseline shift associated with covert spatial attention, many studies analyzing spectral elements of the EEG do find such evidence. Most prominently, covertly directing spatial attention to one visual field is known to produce a decrease in alpha-band power in the contralateral hemisphere and an increase in alpha-band power in the ipsilateral hemisphere (cf. *Research Spotlight 3.2*). An ERP analysis would not detect such an effect,

because alpha-band dynamics represent a major component of the “noise” that ERP averaging is intended to minimize (see *Methodology Box 4.2*). These considerations also highlight the fact that, for some questions, to “really know what’s going on,” one needs to get inside the cranium.

Effects of attention on neuronal activity

Spatial attention

Figure 6.19 illustrates a study performed in the laboratory of Robert Desimone, the neurophysiologist whose work we first encountered in *Chapter 5*, when, as a trainee of Charles Gross, he studied the visual functions of IT cortex. Subsequently, Desimone joined the fabled Laboratory for Neuropsychology at the National Institutes of Health, where he and his trainees carried out many pioneering studies on the neurophysiological bases of visual attention

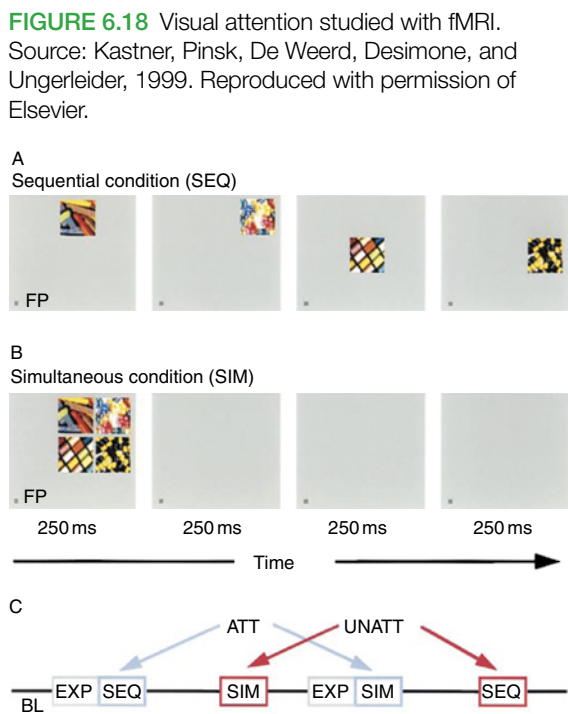
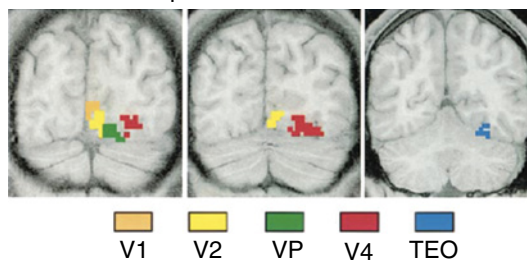


FIGURE 6.18.A Rows **A** and **B** illustrate one trial of the two stimulus presentation conditions. Each 10 second block (represented as rectangles labeled “SEQ” and “SIM” in timeline of row **C**) included 10 trials during which four stimuli were presented either sequentially or simultaneously, always to the upper right visual field. During “UNATT”(-ended) blocks subjects ignored these, and during “ATT”(-ended) blocks they counted the number of occurrences of a pre-identified image. ATT blocks were preceded by a 10 second “EXP” (-ectation) period during which subjects covertly attended to the location where the stimuli would appear. (FP=fixation point.)

A Unattended presentations versus control



B Expectation versus control

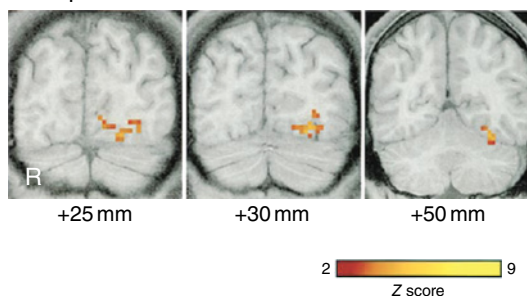


FIGURE 6.18.B Panel **A** shows regions of the left hemisphere with elevated fMRI activity during “UNATT” vs. “BL.” Panel **B** shows regions with elevated activity during “EXP” vs. “BL.”

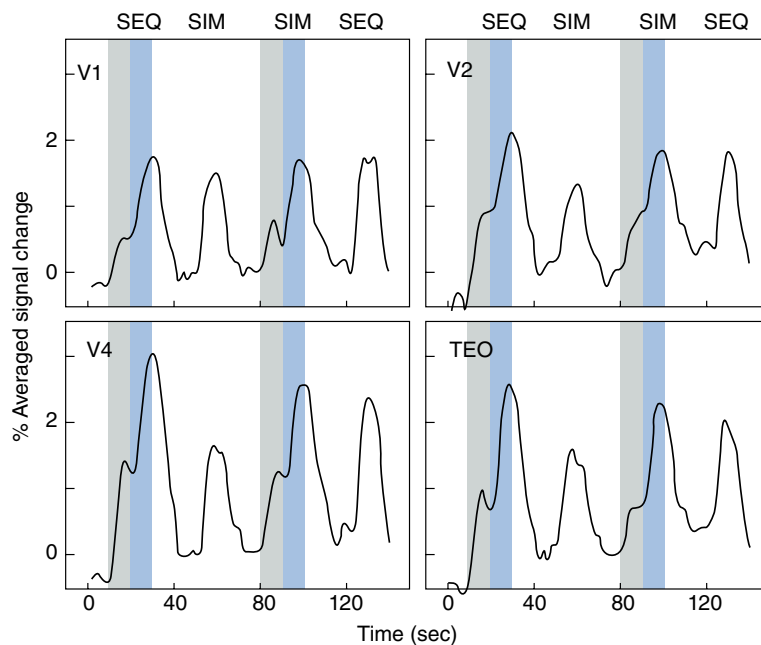


FIGURE 6.18.C Time courses of fMRI activity in the activated voxels from *Figure 6.18.B*. The labels “SEQ” and “SIM” indicate when stimuli were presented, the gray band highlighting the EXP period, the blue band ATT blocks, and “unbanded” regions unattended portions of the scan (including BL).

FIGURE 6.19 The effects of spatial attention on neural activity in the macaque visual system. Luck, Chelazzi, Hillyard, and Desimone, 1997. Reproduced with permission of The American Physiological Society (APS).

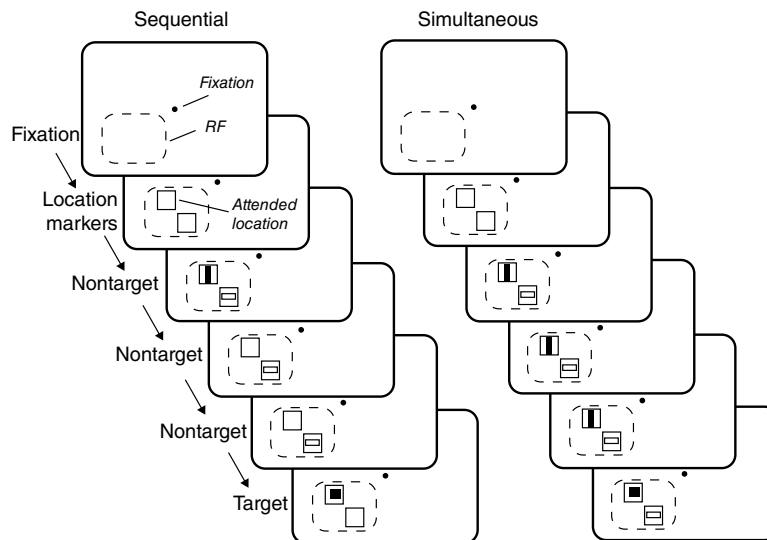


FIGURE 6.19.A The behavioral task, in which stimuli were presented sequentially or simultaneously in two “marked” locations. In different conditions, these locations were either both placed within the receptive field of the neuron in V2 or V4 that was being recorded from, or one was inside and one outside the receptive field. At the beginning of each block of trials, the animal was instructed which of the two marked locations was to be attended. Of interest here is activity from neurons in conditions when one of the marked locations was located outside the receptive field (not shown here).

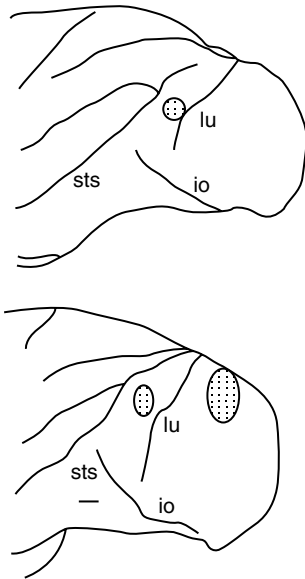


FIGURE 6.19.B Locations of electrode penetration and recording sites in two monkeys. V4 recordings were made within shaded oval regions anterior to the lunate sulcus (lu) in both animals. In the animal illustrated below, recordings were also made in V1 (shaded oval posterior to the lunate sulcus) and in V2, reached through penetrations extending rostrally from the V1 recording site to the caudal bank of lu, where V2 is located. io = inferior occipital sulcus; sts = superior temporal sulcus.

Tip 6.9

The training “genealogy” of many neuroscientists can be found at www.neurotree.org.

(*Tip 6.9*). As with studies that we’ve already considered, the subject (in this case, a monkey) covertly attended a specific location and waited for a pre-designated target to appear there. Recordings from neurons in areas V2 and V4 clearly showed that covertly attending to a location within their receptive fields produced a sustained increase in their firing rate – a baseline shift. From these findings, the authors concluded that “visual-spatial attention operates, at least in part, by creating a preset sensory bias that modulates the initial volley of sensory information

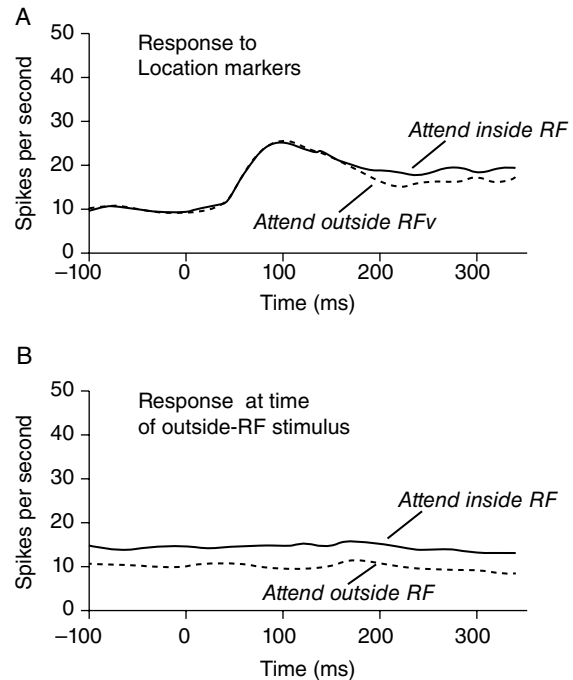


FIGURE 6.19.C Two examples of a “baseline shift” effect of spatial attention in the average firing rate of 40 V4 neurons. **A** illustrates the prestimulus period (labeled “Location markers” in panel A): onset of location markers occurred at time 0, and approximately 175 ms later the neurons’ firing rates diverged depending on whether attention was focused inside or outside the receptive field (RF). **B** illustrates the middle of the trial: a nontarget stimulus is presented outside the RF at time 0, when the animal is attending to a marked location inside vs. outside the RF.

as it passes through area V4” (Luck, Chelazzi, Hillyard, and Desimone, 1997, p. 41).

At this point, the reader may be wondering why the studies illustrated in *Figure 6.18* and *Figure 6.19* each included both simultaneous and sequential stimulus presentation procedures. The answer will become clear in the next, and final, subsection of this chapter.

Object-based attention

As we near the end of this chapter we turn to a final theme invoked by the study of unilateral neglect – the *competition* for stimulus representation inferred from the phenomenon of extinction. Evidence for such

competition at the neuronal level has come from research from Desimone's group (with collaboration from the British experimental psychologist John Duncan, who visited Desimone's lab while on leave from the Medical Research Council Applied Psychology Unit in Cambridge, UK), illustrated in *Figure 6.20*. The first thing to note is that these recordings were made in vAIT, where, as we saw in the previous chapter, neurons show a high degree of preference for the conjunctions of features that define specific objects. The first 150 ms after array onset shows the evidence for competition between the neural representations of objects: the response to a "good" target stimulus is suppressed when it is paired with a second stimulus ("Target = good stim. in 2-stim. array") relative to when it is presented alone ("Target = good stimulus alone"). This is presumably due to **lateral inhibition** that occurs when collateral branches of neuron A's axon synapse on inhibitory interneurons located in cortical columns that are

tuned for different stimulus configurations, and these interneurons, in turn, release GABA to suppress activity within that column. That is, the "initial [feed-forward] volley of sensory information" associated with each of the objects in the two-item array results in the neural representations of each acting to suppress the neural representation of the other. The influence of attention is seen at the inflection, at ~175 ms, of the response to the target when it is the "good stim. in 2-stim. array" (as labeled in *Figure 6.20*). From that point forward, the neuron's response is identical in the "alone" and "2-stim" conditions – the suppression from the competing stimulus representation has been overcome. In the parlance of Desimone and Duncan and colleagues, the *competition* has been *biased* by the top-down influence of selective attention.

The competition is also seen in responses when the "poor" stimulus is the target. When presented alone, the poor stimulus, by definition, barely influences the activity of the neuron. When the neuron's

FIGURE 6.20 Neural evidence for biased competition as a mechanism of object-based attention. Source: Chelazzi, Duncan, Miller, and Desimone, 1998. Reproduced with permission of The American Physiological Society (APS).

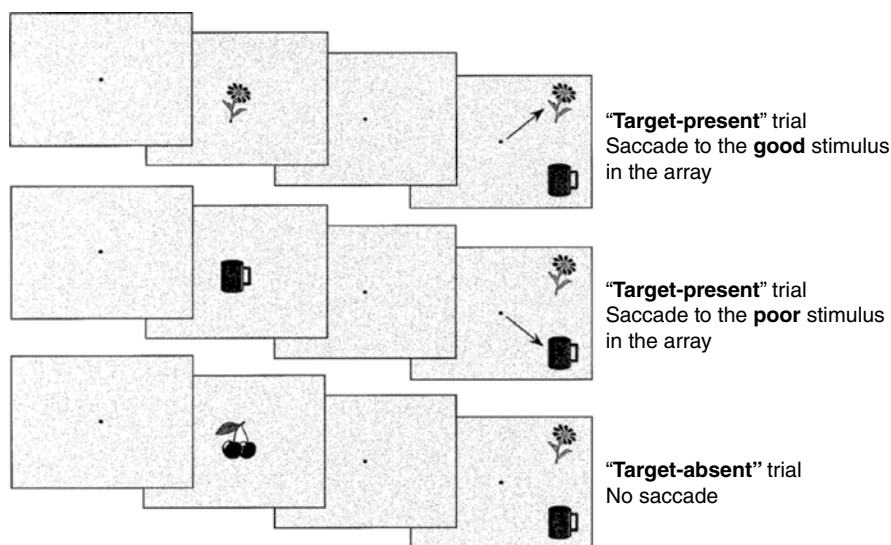


FIGURE 6.20.A The search task (*Tip 6.10*). The target object is presented at the beginning of the trial, and after the delay, an array of two stimuli is presented and the monkey makes a saccade to the target (no saccade if the target is not in the array). Not shown are trials with 1-item arrays, which could also be "target-present" or "target absent." Once a neuron was isolated, the experimenters first determined what stimuli were effective at driving it and what were not. "Effectiveness" of target object (good, bad) was then used as an independent variable in the task.

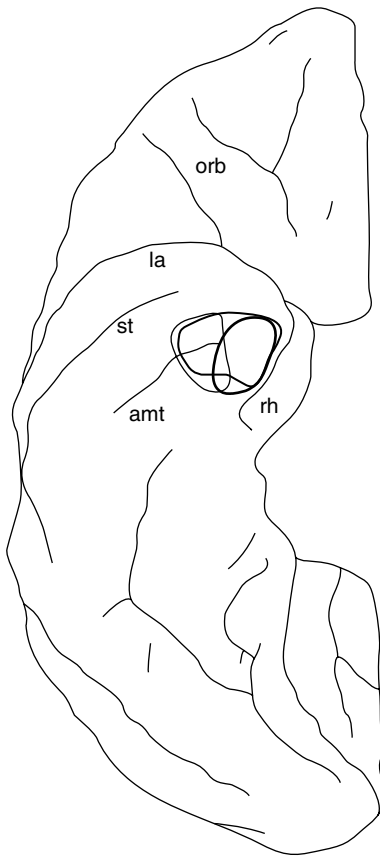


FIGURE 6.20.B The recording locations in vAIT of three monkeys, superimposed on one image of the ventral surface of a right hemisphere, with the occipital pole at the bottom of the image, the frontal pole at the top, and the lateral surface on the left. The labeled sulci are the rhinal (rh), anterior middle temporal (amt), superior temporal (st), lateral (la), and orbitofrontal (orb).

good stimulus is presented along with the poor target, however, the neuron initially responds to its preferred stimulus. After ~125 ms, however, the neuron's response to its preferred stimulus begins to be tamped down, presumably because the neurons representing the target have begun winning the competition, and are suppressing the neural representation of the non-target stimulus.

These results from Chelazzi et al. (1998) suggest a neural mechanism that implements the psychological process invoked by James 100 years before: “[Attention] implies withdrawal from some things in order to deal

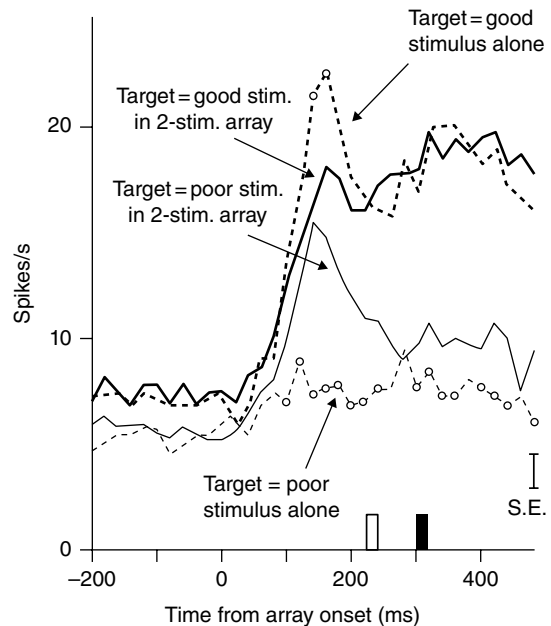


FIGURE 6.20.C A summary of the results from 58 neurons, contrasting good vs. bad targets in 2 stimulus-array and 1 stimulus-array (“alone”) trials. Neural activity is time-locked to the onset of the search array (i.e., array appeared at time 0). Circles indicated time points for which the two good traces or the two poor traces differ statistically from one another. Unfilled and filled bars on horizontal axis indicate average response times to 1-item and 2-item arrays, respectively.

effectively with others” (James, 1950, p. 404). A topic to be taken up in *Chapter 8* is “What is the source of this biasing influence?”

TURNING OUR ATTENTION TO THE FUTURE

In this chapter we started by considering how aspects of behavior that many of us take for granted (e.g., being aware that there’s food on both the right *and* the left side of our dinner plate) can be severely disrupted as a result of a stroke affecting temporoparietal cortical regions, and ended with a consideration of neural phenomena that may underlie some of these behaviors. In the next two chapters we’ll build on what we’ve learned about the properties of the

Tip 6.10

Task naming is often influenced by theoretical context. The authors (correctly) characterized this as a search task, in reference to the visual search task that has a long history in the psychological study of attention. “Delayed match-to-sample” (DMS) would be equally valid, and more informative about what the animals actually had to do. DMS, however, is more strongly associated with long-term memory research (*Chapter 10*).

parietal cortex and of spatial cognition to introduce several new ideas about the neural organization of behavior and cognition. One, already hinted at on

more than one occasion, is that an alternative way to construe the functions of the dorsal stream is that it governs *how* the body can interact with objects that necessarily occupy different locations in the environment. A second will be how parietal and frontal cortical systems coordinate with subcortical systems to produce the finely controlled actions that we execute countless times each day. A third will be to consider the source(s) of attentional control. These include the ideas that PPC and other regions support a dynamic “salience map” that guides our perception of the world, that the volitional vs. the reflexive control of attention are governed by different neural systems.

At a simplistic, yet profound, level, we are about to make our first foray across the central sulcus, into the frontal lobe.

END-OF-CHAPTER QUESTIONS

1. What symptom(s) of unilateral neglect inform ideas about: the lateralization of attention? how we process “near” vs. “far” space? the competition for representation within the brain? location-based vs. object-based attention?
2. In what ways have neuroimaging studies supported, and extended, neuropsychology-based ideas about the functions of the dorsal visual processing stream?
3. We saw how the integration of retinotopic sensory information with eye-position information results in a coordinate transformation from retina-centered to head-centered coordinates. Using the same logic, how could the brain effect a coordinate transformation from head-centered to trunk-centered coordinates?
4. What biophysical and neurophysiological principles underlie PET imaging? fMRI? What are some of the important differences between these two imaging modalities?
5. Summarize the idea of a coordinate transformation. Why is this a critical principle underlying the guidance of action with perception? What evidence is there for the computation of coordinate transformations in the PPC?
6. What is the difference between an egocentric vs. an allocentric representation of space? Summarize neurophysiological evidence that both are maintained in the brain.
7. In what ways is the analogy to a spotlight appropriate for understanding visual attention? Describe (at least) two aspects of attention that do not conform to this analogy.
8. How do location-based and object- (or feature-) based attention differ neurophysiologically?
9. Summarize the evidence that biased competition may be an important principle for visual attention.

REFERENCES

- Andersen, Richard A., Greg K. Essick, and Ralph M. Siegel. 1985. “Encoding of Spatial Location by Posterior Parietal Neurons.” *Science* 230 (4724): 456–458.
- Anllo-Vento, Lourdes, and Steven A. Hillyard. 1996. “Selective Attention to the Color and Direction of Moving Stimuli: Electrophysiological Correlates of Hierarchical Feature Selection.” *Perception & Psychophysics* 58 (2): 191–206. doi: 10.3758/BF03211875.
- Bisiach, Edoardo, and Claudio Luzzatti. 1978. “Unilateral Neglect of Representational Space.” *Cortex* 14 (1): 129–133.

- Burwell, Rebecca D., and Kara L. Agster. 2008. "Anatomy of the Hippocampus and the Declarative Memory System." In *Memory Systems*, edited by Howard Eichenbaum. Vol. 3 of *Learning and Memory: A Comprehensive Reference*, edited by John H. Byrne. 4 vols., 47–66. Oxford: Elsevier.
- Chatterjee, Anjan. 2003. "Neglect: A Disorder of Spatial Attention." In *Neurological Foundations of Cognitive Neuroscience*, edited by Mark D'Esposito, 1–26. Cambridge, MA: MIT Press.
- Chelazzi, Leonardo, John Duncan, Earl K. Miller, and Robert Desimone. 1998. "Responses of Neurons in Inferior Temporal Cortex During Memory-Guided Visual Search." *Journal of Neurophysiology* 80 (6): 2918–2940.
- Di Russo, Francesco, Antígona Martínez, and Steven A. Hillyard. 2003. "Source Analysis of Event-Related Cortical Activity During Visuo-Spatial Attention." *Cerebral Cortex* 13 (5): 486–499. doi: 10.1093/cercor/13.5.486.
- Driver, Jon, Patrik Vuilleumier, and Masud Husain. 2004. "Spatial Neglect and Extinction." In *The Cognitive Neurosciences III*, edited by Michael S. Gazzaniga, 589–606. Cambridge, MA: MIT Press.
- Haxby, James V., Cheryl L. Grady, Barry Horwitz, Leslie G. Ungerleider, Mortimer Mishkin, Richard E. Carson, ... Stanley I. Rapoport. 1991. "Dissociation of Object and Spatial Visual Processing Pathways in Human Extrastriate Cortex." *Proceedings of the National Academy of Sciences of the United States of America* 88 (5): 1621–1625. doi: 10.1073/pnas.88.5.1621.
- James, William. 1950. *The Principles of Psychology*, vol. 1. New York: Dover. First published 1890.
- Karnath, Hans-Otto, Johannes Rennig, Leif Johannsen, and Chris Rorden. 2011. "The Anatomy Underlying Acute Versus Chronic Spatial Neglect: A Longitudinal Study." *Brain* 134 (3): 903–912. doi: 10.1093/brain/awq355.
- Kastner, Sabine, Mark A. Pinsk, Peter De Weerd, Robert Desimone, and Leslie G. Ungerleider. 1999. "Increased Activity in Human Visual Cortex During Directed Attention in the Absence of Visual Stimulation." *Neuron* 22 (4): 751–761. doi: 10.1016/S0896-6273(00)80734-5.
- Luck, Steven J., Leonardo Chelazzi, Steven A. Hillyard, and Robert Desimone. 1997. "Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex." *Journal of Neurophysiology* 77 (1): 24–42.
- Menon, Ravi S., and Seong-Gi Kim. 1999. "Spatial and Temporal Limits in Cognitive Neuroimaging with fMRI." *Trends in Cognitive Sciences* 3 (6): 207–215. doi: 10.1016/S1364-6613(99)01329-7.
- O'Keefe, John, and Lynn Nadel. 1978. *The Hippocampus As a Cognitive Map*. Oxford: Oxford University Press.
- Pashler, Harold E. (1999). *The Psychology of Attention*. Cambridge, MA: MIT Press.
- Petersen, Steven E., Peter T. Fox, Michael I. Posner, Mark Mintun, and Marcus E. Raichle. 1988. "Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing." *Nature* 331 (6157): 585–589. doi: 10.1038/331585a0.
- Posner, Michael I. 1980. "Orienting of Attention." *Quarterly Journal of Experimental Psychology* 32 (1): 3–25. doi: 10.1080/00335558008248231.
- Raichle, Marcus E. 2010. "Two Views of Brain Function." *Trends in Cognitive Neuroscience* 14 (4): 180–190. doi: 10.1016/j.tics.2010.01.008.
- Saygin, Ayse Pinar, and Martin I. Sereno. 2008. "Retinotopy and Attention in Human Occipital, Temporal, Parietal, and Frontal Cortex." *Cerebral Cortex* 18 (9): 2158–2168. doi: 10.1093/cercor/bhm242.
- Snyder, Larry H., Aaron P. Batista, and Richard A. Andersen. 1997. "Coding of Intention in the Posterior Parietal Cortex." *Nature* 386 (6621): 167–170. doi: 10.1038/386167a0.
- Thier, Peter, and Richard A. Andersen. 1998. "Electrical Microstimulation Distinguishes Distinct Saccade-Related Areas in the Posterior Parietal Cortex." *Journal of Neurophysiology* 80 (4): 1713–1735.

OTHER SOURCES USED

- Aguirre, Geoffrey K., Eric Zarahn, and Mark D'Esposito. 1998. "The Variability of Human, BOLD Hemodynamic Responses." *NeuroImage* 8 (4): 360–369. doi: 10.1006/nimg.1998.0369.
- Burgess, Neil. 2006. "Spatial Memory: How Egocentric and Allocentric Combine." *Trends in Cognitive Sciences* 10 (12): 551–557. doi: 10.1016/j.tics.2006.10.005.
- Halligan, Peter W., and John C. Marshall. 1991. "Left Neglect for Near but not Far Space in Man." *Nature* 350 (6318): 498–500. doi: 10.1038/350498a0.
- Villringer, Arno. 2000. "Physiological Changes During Brain Activation." In *Functional MRI*, edited by C. T. W. Moonen and Peter A. Bandettini. Heidelberg: Springer-Verlag.
- Weintraub, Sandra, and M.-Marsel Mesulam. 1989. "Neglect: Hemispheric Specialization, Behavioral Components and Anatomical Correlates." In *Handbook of Neuropsychology*, vol. 2, edited by Francois Boller and Jordan Grafman, 357–374. Amsterdam: Elsevier.
- Wilson, Matthew A., and Bruce L. McNaughton. 1994. "Reactivation of Hippocampal Ensemble Memories During Sleep." *Science* 265 (5172): 676–679. doi: 10.1126/science.8036517.

FURTHER READING

- Aguirre, Geoffrey K. 2003. "Topographical Disorientation: A Disorder of Way-Finding Ability." In *Neurological Foundations of Cognitive Neuroscience*, edited by Mark D'Esposito, 90–108. Cambridge, MA: MIT Press.
Thorough overview of the clinical aspects of this intriguing disorder, and its implications for our understanding of spatial cognition.
- Burgess, Neil. 2006. "Spatial Memory: How Egocentric and Allocentric Combine." *Trends in Cognitive Science* 10 (12): 551–557. doi: 10.1016/j.tics.2006.10.005.
Synthesis of human behavioral studies of navigation and spatial representation with the neurophysiological representation of space in the rat, monkey, and human to argue that parallel, interacting ego- and allocentric reference frames guide spatial cognition.
- Desimone, Robert, and John Duncan. 1995. "Neural Mechanisms of Selective Visual Attention." *Annual Review of Neuroscience* 18:193–222. doi: 10.1146/annurev.ne.18.030195.001205.
The highly influential articulation of the principle of biased competition that draws equally from theory, experimental psychology, and neurophysiology.
- Lezak, Muriel D. 1995. *Neuropsychological Assessment*, 3rd ed. New York: Oxford University Press.
This is the authoritative compendium of (seemingly) every neuropsychological test ever developed, how it is administered, and what behavioral and/or cognitive functions that test measures. One could use it to look up, for example, the Benton Facial Recognition Test.
- O'Leary, Maureen A., Jonathan I. Bloch, John J. Flynn, Timothy J. Gaudin, Andres Giallombardo, Norberto P. Giannini, ... Andrea L. Cirranello. 2013. "The Placental Mammal Ancestor and the Post-K-Pg Radiation of Placentals." *Science* 339 (6120): 662–667. doi: 10.1126/science.1229237.
Description of evidence for a "hypothetical placental ancestor" proposed to be a progenitor species of all currently extant placental mammals (including rodents and primates); Figure 2 includes an artist's impression of what it may have looked like, the inferred structure of its brain, and of its ear ossicles and ectotympanic bone.

CHAPTER 7

SKELETOMOTOR CONTROL

KEY THEMES

- Motor control is governed by three principal circuits: the corticospinal system that carries efferent motor commands from motor cortex to the spinal circuits that carry out motor commands; and the cortico-cerebellar and cortico-basal ganglia systems, each of which processes copies of the efferent signal and feeds back to motor cortex so as to refine and/or facilitate future actions.
- The motor cortex plans and generates commands in abstract x,y,z coordinates, and the spinal cord translates these commands (often nonlinearly) into the set of muscle tensions needed to carry out the action.
- Despite its generally homuncular functional topography, primary motor cortex differs from sensory analogues in that its representation of the body surface is not highly precise, and it also represents actions that are not tied to any one part of the body.
- M1 neurons have relatively broad tuning, and precise control of movement is achieved via population coding – the summation of input from thousands of units firing at varying intensities.
- Electrophysiological and microstimulation evidence that PPC neurons contribute to planning and execution of eye movements and reaches, combined with neuropsychological evidence, have been taken as evidence that a core function of PPC may be to determine *how* visual information can be used to guide behavior.
- The unique wiring of the cerebellum makes it a key neural substrate for motor learning, fine-tuning ongoing movements, and, perhaps, encoding and refining cognitive models of the world.
- The cortico-basal ganglia-thalamic circuits integrate information from many cortical territories and use the output to gate frontal cortex activity.
- Common principles of network and neurochemical dynamics (including a key role for dopamine) across “affective,” “cognitive,” and “motoric” basal ganglia systems suggest that many principles of motor control generalize to these other domains of behavior.
- The discovery of mirror neurons in premotor cortex has generated intriguing theories ranging from the understanding of actions and intentions of others, to atypical neurological development (e.g., autism spectrum disorder), to the evolution of human culture.

CONTENTS

KEY THEMES

THE ORGANIZATION OF THE MOTOR SYSTEM

The anatomy of the motor system
The corticospinal tract
The cortico-cerebellar circuit
The cortico-basal ganglia-thalamic circuits

FUNCTIONAL PRINCIPLES OF MOTOR CONTROL

The biomechanics of motor control
 Muscles behave like springs
 Motor computations in the spinal cord
Motor cortex
 The motor homunculus
The neurophysiology of movement
 EEG and LFP correlates
 Single unit activity reveals population coding

MOTOR CONTROL OUTSIDE OF MOTOR CORTEX

Parietal cortex: guiding *how* we move
Cerebellum: motor learning, balance, ... and mental representation?
 Cerebellar circuitry
 Cerebellar functions
Basal ganglia
 A subcortical gate on cortical activity
 The role of dopamine

COGNITIVE FUNCTIONS OF THE MOTOR SYSTEM

IT'S ALL ABOUT ACTION

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

Our introduction to the frontal lobes came in *Chapter 1*, where we considered a nineteenth-century breakthrough in thinking about how the brain controls movement of the body, and one relating to the production of speech. The former is most relevant here. The observations of the neurologist Jackson gave rise to the idea of a somatotopic organization of the brain system controlling movement in the human. The electrical stimulation experiments of Fritsch and Hitzig indicated that motor cortex, in the dog, was localized to frontal cortex, and that distinct **motor effectors** contralateral to stimulation sites were represented by discrete regions within this lobe. In addition to fleshing out more detail about the structure and function of the motor system, this chapter will introduce many concepts related to the control of action that will be highly germane to many of the remaining topics to be covered in this book, including *cognitive control*, *decision making*, *language*, *social behavior*, and *consciousness*. Indeed, some have argued that much of our thought and behavior, no matter how abstract and “high-level,” can be construed as more-or-less abstract implementations of the same principles of motor control that we will consider in detail in this chapter.

THE ORGANIZATION OF THE MOTOR SYSTEM

The anatomy of the motor system

Primary motor cortex (M1) occupies the anterior bank of the central sulcus and the caudal portion of the precentral gyrus, a region classified as BA 4 that stretches from the medial wall of the hemisphere to the dorsal lip of the Sylvian fissure. In loose analogy to cortical sensory systems, there are two adjacent motor-related fields with “second-level functions,” both sharing BA 6: the premotor cortex (PMC) of the rostral half of the precentral gyrus; and the supplementary motor area (SMA) located superior to PMC and rostral to the superior-most and medial portions of M1. Unlike with sensory systems, however, the functions of M1, PMC, and SMA aren’t as hierarchically organized as are, say, V1, V2, and MT. For example, each of these three motor areas sends projections from cortex to the areas of the spinal cord that

directly control movement. (Thus, for simplicity, when describing general properties that apply to each, the term “motor cortex” will be used to refer nonspecifically to M1, PMC, and SMA.)

The motor cortices are a central hub in three circuits that carry action-related signals. The fibers carrying the signals that descend along the ventral spinal cord to trigger movements of the body are called **motor efferents**. The other two carry copies of these motor signals, but are considered **recurrent circuits** (often called “loops”) in that they eventually feed back onto the frontal cortex. They implement the principle of **efference copy**, whereby a copy of the efferent signal can be used to, for example, compare the intended outcome of the outgoing motor command against information about the actual outcome (conveyed by ascending sensory signals) and, if the two don’t match up, send a corrective signal back to the motor cortex so that the next movement will be more successful. Let’s consider each circuit in turn.

The corticospinal tract

The giant pyramidal cells of layer V of motor cortex send axons that descend through the brain, funneling together as the internal capsule that passes between the (laterally located) lentiform nucleus of the basal ganglia and the (medially located) caudate nucleus and thalamus (see *Figure 2.4* and *Web Link 1.1*), decussating in the caudal brain stem, and descending in the ventral spinal cord as the corticospinal tract (*Figure 7.1*). Each corticospinal fiber activates motor neurons in the cord either via direct synaptic connections or, most typically, indirectly via local circuitry in the cord. Motor neurons send axons that innervate muscles, but because their cell bodies are located within the spinal cord, they are considered part of the CNS. Their functional organization can be considered from two perspectives: from that of the muscle, each muscle is innervated by a grouping of neurons called a motor neuron pool; from that of the individual motor neuron, it and each of the motor fibers that it innervates make up a motor unit. As we saw in *Real-World Applications 2.2*, motor neurons release the ACh at their synapses, which are often referred to as the neuromuscular junction. More on muscles in the upcoming section on *The biomechanics of motor control*.

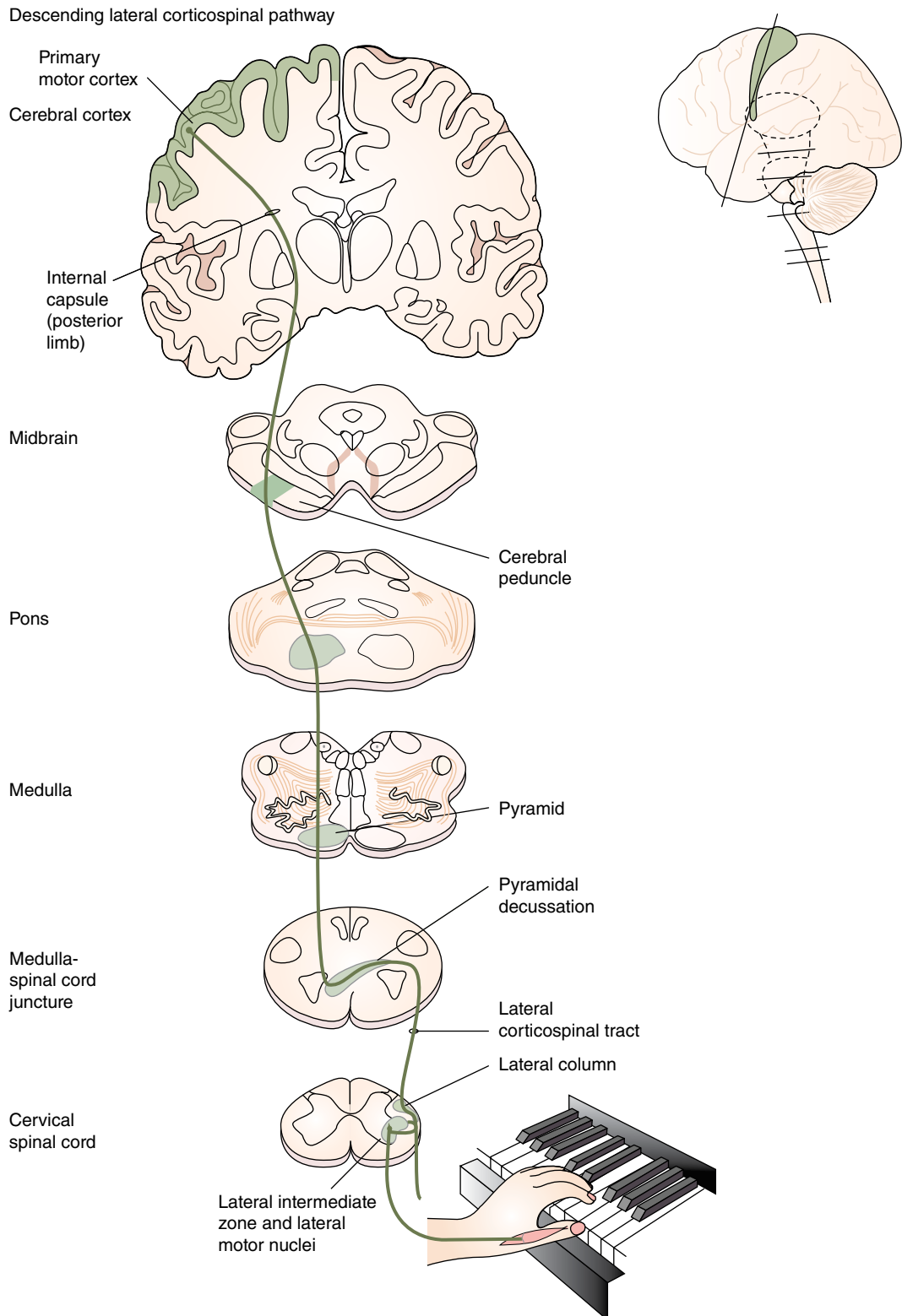


FIGURE 7.1 The descending corticospinal pathway. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

The cortico-cerebellar circuit

The cerebellum is an intriguing structure, in that it contains 50 billion of the roughly 80 billion neurons of the brain, yet, as the neurobiologist and theorist Giulio Tononi has noted, its complete removal has effectively no impact on the subjective sense of consciousness. It is made up of hundreds of thousands of parallel, modular (i.e., not interconnected) circuits that are constantly comparing motor output to sensory feedback so as to fine-tune the targeting and smooth execution of our movements, and to help us maintain postural stability. The circuit is one way, linking frontal cortex to pons to cerebellar cortex to cerebellar deep nuclei to thalamus and back to frontal cortex. The densely foliated cerebellar cortex is shaped like one half of a sawed-in-two tennis ball, but with thicker walls due to deep folds between the lobes of the cerebellar cortex, and the fibers projecting from this cortex to the deep nuclei lying near the “deepest” point of this concavity. Also projecting into the cerebellum are ascending projections carrying information about the periphery (touch on the skin, body configuration from muscle stretch receptors and joint-position receptors) and overall body position and balance from the vestibular system. The workings and functions of the cerebellum will be considered the upcoming section on motor learning.

The cortico-basal ganglia-thalamic circuits

This circuit and its nuclei play important roles in motor control, learning, and motivated behavior, and a variety of dysfunctions within it can result in disorders ranging from Parkinson’s disease (impairing movement and high-level cognition) to unilateral neglect to pathological gambling. As illustrated in *Figure 7.2*, its unidirectional circuits comprise a “direct pathway” (cortex to neostriatum to internal segment of the globus pallidus (GPi)/substantia nigra pars reticulata (SNpr) to thalamus back to cortex) and an “indirect pathway” (cortex to neostriatum to external segment of the globus pallidus (GPe) to subthalamic nucleus to GPi/SNpr to thalamus back to cortex). At the gross anatomical level, there are five relatively segregated *cortico-basal ganglia-thalamic*

Tip 7.1

There is evidence for smaller projections of some of these circuits back to non-frontal cortical regions (e.g., back to IT in the ventrolateral PFC loop), a level of detail that we won’t consider further.

Tip 7.2

The distinctive black pigmentation of the substantia nigra (“black substance” in Latin), visible to the naked eye, derives from the high concentration in its neurons of the enzyme tyrosine hydroxylase, which is critical for the synthesis of dopamine.

circuits, each associated with dissociable classes of behavior: *motor*, *oculomotor*, *dorsolateral prefrontal*, *lateral orbitofrontal*, and *limbic* (*Figure 7.3.A*). *Figure 7.3.B* illustrates the connectivity of the cortico-basal ganglia-thalamic motor circuit (for details on the others see Alexander, DeLong, and Strick, 1986; Alexander and Crutcher, 1990). Although the focus during this chapter will be the motor circuit, a noteworthy property of each of these circuits is the way each integrates inputs from multiple cortical regions and funnels the integrated result back onto the originating region of frontal cortex (*Tip 7.1*). Another common feature of these circuits is that the neostriatal node of each is densely innervated by dopaminergic fibers from a midbrain nucleus called the substantia nigra pars compacta (SNpc; *Tip 7.2*).

FUNCTIONAL PRINCIPLES OF MOTOR CONTROL

The biomechanics of motor control

Although moving the body in a coordinated fashion is something that we do effortlessly, often “thoughtlessly,” this subjective facility masks a very complicated engineering problem. To summarize just one example, the movements of our arms from almost any points

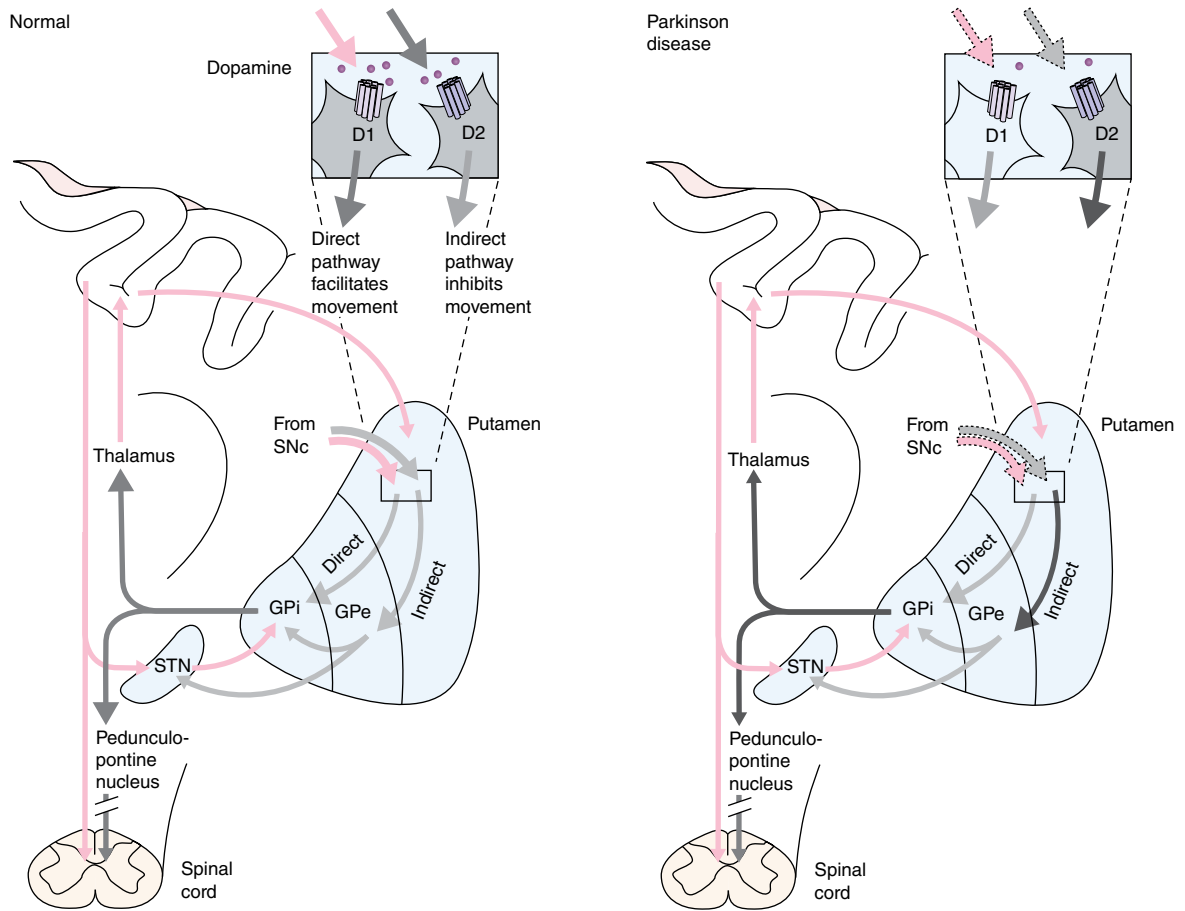


FIGURE 7.2 Schematic illustrations of the direct and indirect pathways of the basal ganglia, in the healthy (“Normal”) state on the left, and in Parkinson’s disease on the right. Red arrows connote excitatory pathways, gray inhibitory, and the change of saturation in the Parkinson’s disease figure whether that pathway strengthens or weakens as a result of disease processes. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

A to B are remarkably straight, this despite the fact that executing such movements entails the precisely coordinated, simultaneous changing of joint angles – at the shoulder and the elbow (*Figure 7.4*). How does the nervous system accomplish this?

Muscles behave like springs

The tissue that puts the “motor” in “motor control,” of course, is muscle. To move one’s hand from a point A to a point B, it is necessary to contract the appropriate sets of muscles such that the joints that connect the bones of the arm move in the ways illustrated in

the lower columns of *Figure 7.4*. Note that at this most proximal level of motor control, any one muscle can only influence a single joint angle. For example, the antagonistically paired biceps and triceps muscles can change the angle of the elbow – nothing more. Thus, to execute even a simple “hand movement” as in *Figure 7.4*, the appropriate neural commands have to be delivered to at least four sets of muscles (biceps and triceps for the elbow, the analogous muscles for the shoulder) in a precise, exquisitely choreographed sequence that will produce a smooth, straight motion. Note from this figure that movements to the same

FIGURE 7.3 The five cortico-basal ganglia-thalamic circuits. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

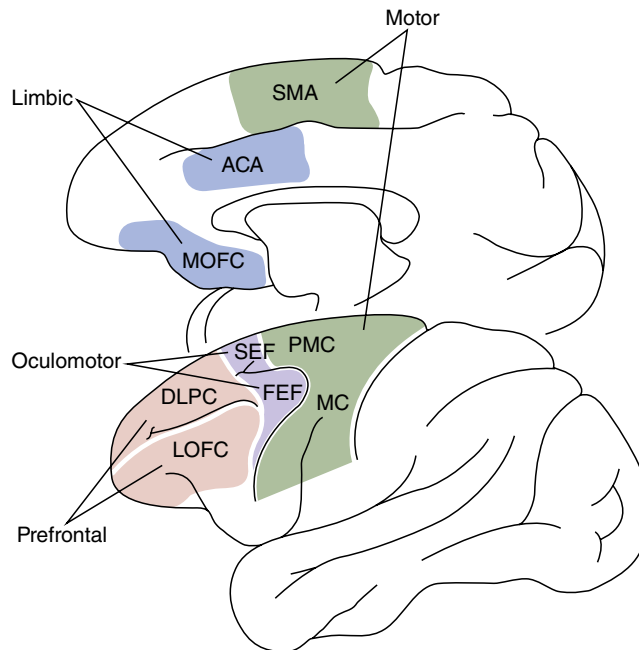


FIGURE 7.3.A Frontal targets the five loops. ACA = anterior cingulate area; MOFC = medial orbital frontal cortex; LOFC = lateral orbitofrontal cortex.

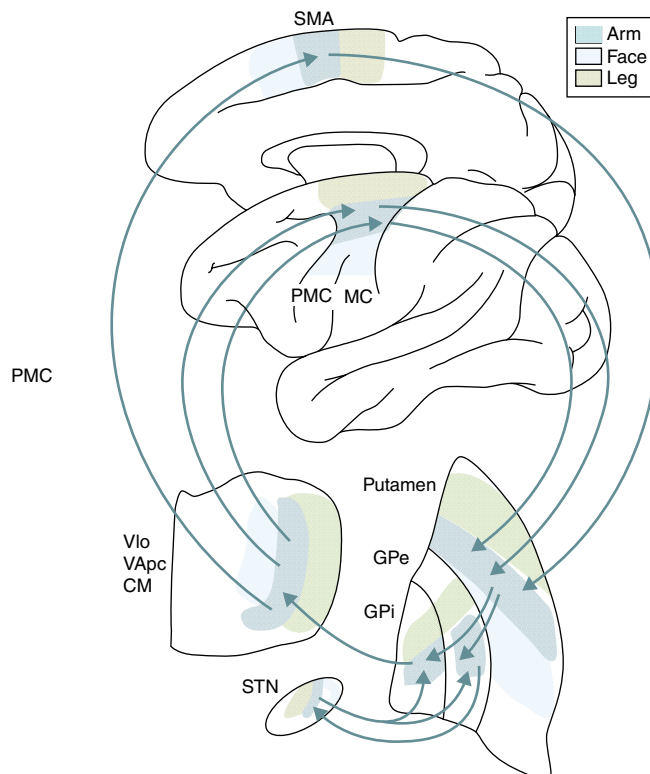


FIGURE 7.3.B Diagram of segregation of somatotopic information at each stage of the motor circuit.

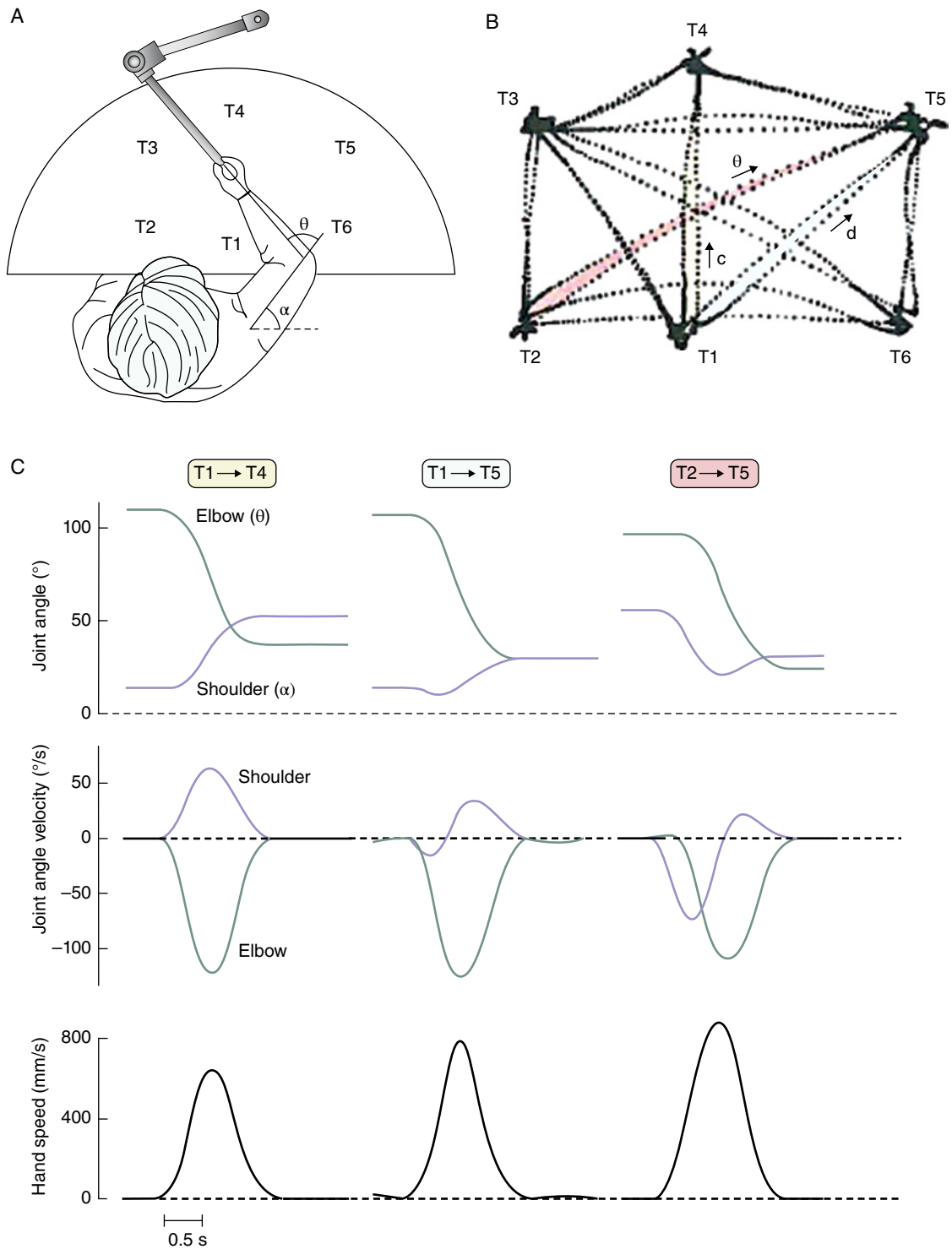


FIGURE 7.4 **A.** Top-down view of a subject grasping the handle of a manipulandum, which can be moved in a 2-D plane to each of the six targets indicated on the table. **B.** Movement paths (two between each target) for a subject. **C.** Joint angle, joint angular velocity (i.e., speed at which joint angle is changing) and hand speed plotted against time to execute each of three movements from **B**. Source: Kandel, Schwartz, and Jessell, 2000. Reproduced with permission of McGraw-Hill.

location, but starting from different locations, require nonlinearly different transformation of joint angles and joint angular velocities.

Fortunately, the mind-bogglingly complicated sets of nonlinear computations required each time we move our bodies is simplified by a felicitous mechanical property of muscles: they behave like springs. That is, one has only to set the tensions on a set of muscles and they will “automatically” adjust to a new equilibrium point, pulling the bones to which they are attached into a new position. Therefore, in principle, each unique location in x,y,z , coordinates within the peripersonal space of an individual has a unique set of muscle tensions associated with it. For example, from where you are currently sitting (or standing; or reclining; or whatever), there is a unique set of tensions required of the muscles controlling your shoulder, elbow, wrist, and index finger in order for you to touch your finger to this period \rightarrow, \leftarrow (Go ahead, do it. No one is watching.) And so, in principle, your cortex only needs to issue the command *Touch that period* and the requisite muscle-tension settings will automatically deliver your index finger to that point. The nonlinear transformations of joint angles that enable your finger to travel from wherever it was to the period that you are now pointing to “just happen” as a consequence of moving from A to B, they don’t have to be explicitly calculated and commanded (*Tip 7.3*).

Motor computations in the spinal cord

If the motor cortex needs only to specify a new endpoint for, say, the index finger, what translates between that specification and the presumed command to *Go to X,Y,Z* and the resetting of muscle tensions? (That is, to execute a movement, each individual muscle only experiences a change in the amount of ACh being released onto it, it can’t “know” anything about

point X,Y,Z .) The answer is that this translation happens in the networks of interneurons in the spinal cord to which most corticospinal fibers project. Evidence for this was first produced in the frog, subsequently in the rat, in an experimental preparation in which all descending skeletomotor and ascending somatosensory connections between spinal cord and brain are severed. Then, microstimulation is applied to the portion of the spinal cord whose motor neurons innervate, say, the left forelimb. The intriguing finding is that stimulation in this single location always brings the forepaw to the same location in space, regardless of what its starting location was. The area being stimulated, therefore, is converting that burst of electrical stimulation into the new tension settings of the muscles controlling the shoulder and elbow joints that bring the limb into that position in space. Stimulate a second spot in the cord, the limb moves to a different location. Further, a property whose importance will become evident further along in this chapter is that the simultaneous stimulation of both spots moves the limb to a third location. Therefore, in effect, the motor cortex issues the general command, and the spinal cord “takes care of the details.”

In addition to being able to translate cortical commands into precise settings of muscle tensions, the spinal cord also contains what are referred to as “pattern generators.” Thus, for example, early to mid-twentieth-century experiments established that if the connections between an animal’s brain and spinal cord are severed, and the animal is suspended in a harness over a treadmill, it will walk with an appropriate gait when the treadmill starts moving. Not only that, its gait adjusts appropriately to treadmill speed, from a trot to a gallop to a walk. The only explanation for such demonstrations is that remarkably sophisticated sensorimotor control circuitry exists within the spinal cord itself.

Tip 7.3

There are details about a strictly endpoint control scheme, which has been implied here, vs. the equilibrium-point control scheme that is closer to what is implemented by the nervous system, that are beyond the scope of this book.

Motor cortex

Okay, so we’ve established that an animal can effectively walk, gallop, and trot, as appropriate, without any input from the brain. So what does the motor cortex do? (Why do we need it?) In a sentence, it controls *volitional* movement of the body (*Tip 7.4*). Let’s take a closer look.

Tip 7.4

If one were brain-dead, it could be that one's legs would start moving if lowered onto a treadmill. However, one would never reach out and point to a period in response to a verbal command.

The motor homunculus

Unless you have just picked up this book for the first time, and opened to this chapter, you won't be surprised to learn that M1 is organized somatotopically

(one of at least five homuncular representations in motor cortex, as is hinted at in *Figure 7.3.B*). *Figure 7.5* is also taken from surgical mapping studies of Penfield and Rasmussen (1950). One important difference between the primary somatosensory and motor cortices, however, is that the latter's map lacks the same high-fidelity correspondence to different parts of the body. This is in part due to the fact that the smallest unit that the motor cortex can represent corresponds to bones, rather than the surface of the skin. For example, it wouldn't make sense for M1 to represent different parts of the forearm in the way that S1 does, with a set of receptive fields that cover its entire surface. Rather, M1 represents the forearm as a

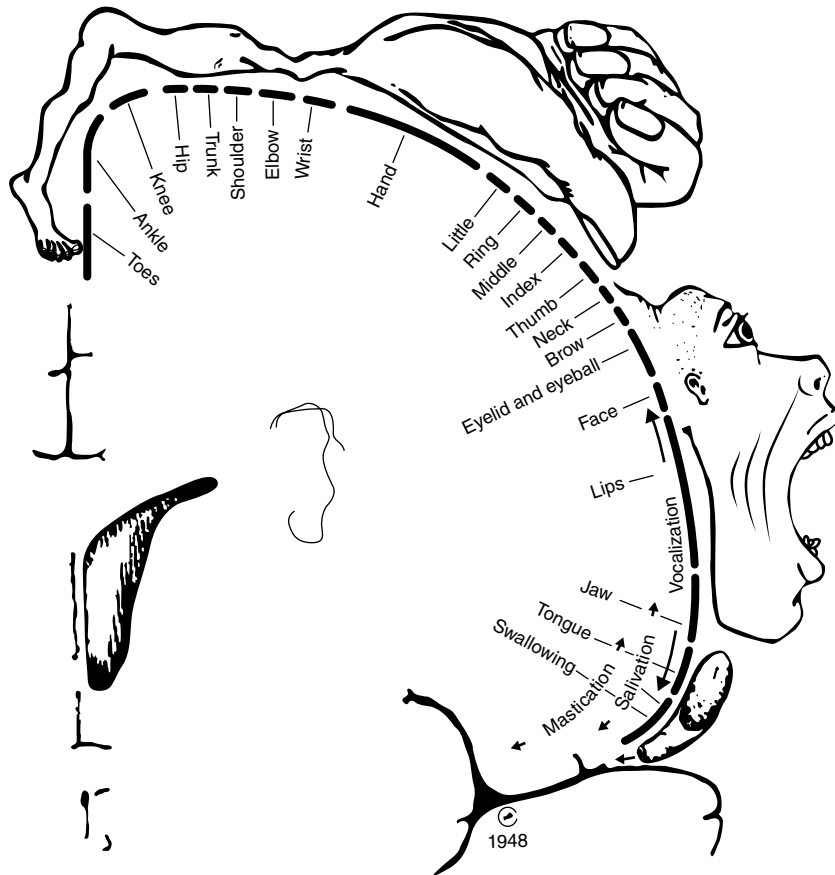


FIGURE 7.5 Penfield and Rassmussen (1950) motor homunculus. Same graphical conventions as *Figure 4.11*. For a more contemporary view of the functional organization of the motor cortex, see *Further Reading* by Kaas (2012). Source: Penfield and Rassmussen, 1950. Reproduced with permission of Macmillan.

unit that can be moved. Another, more fundamental, reason is that M1 represents *actions*, not just parts of the body. For example, it is often pointed out that a person's handwriting looks similar (albeit progressively sloppier) when the pen is held in the right hand vs. the left hand vs. in the mouth vs. between the toes. Thus the representation of the shape of one's handwriting must exist in a format that is accessible by any set of effectors that needs to implement it. Nonetheless, the fact that there is a mapping between different parts of M1 and different parts of the body means that motor cortex doesn't only employ purely abstract representations like "point X,Y,Z" or the shape of one's handwriting. Rather, it clearly chooses which effector(s) will be used to carry out any particular action. Further, it also controls the **kinematics** and **dynamics** of movement, such as the speed with which the arm will move, or the tightness with which the fingers will grip.

How do PMC and SMA differ from M1? For now, we'll content ourselves with the short-hand summary that they contribute to high-level aspects of motor control, such as planning to make a movement, imagining making a movement, and coordinating motor sequences made up of many individual actions. So that's an overview of what motor cortex does. Now, how does it work?

The neurophysiology of movement

EEG and LFP correlates

In the decade following Berger's discovery, and Adrian's confirmation, of the alpha rhythm, one of several cortical rhythms to be discovered was the mu. Although the mu rhythm occupied the same frequency band as alpha, it differed in morphology – as a series of thin, spiky peaks surrounded by rounded troughs (suggesting the shape of the greek letter " μ ") – in topography – being maximal over central electrodes – and in function – being insensitive to perceptual factors, but related to movement. Specifically, the amplitude of the mu rhythm is maximal when the body is at rest, and begins to decline roughly two seconds prior to a movement. Such "mu desynchronizations" can be local to the cortical representation of the effector being moved, if one moves, for example, just the finger, toe, or tongue. Thus, the mu rhythm is analogous to alpha in that it

reflects a state of suppression and/or "idling" in the thalamocortical system producing it.

In the 1950s, the French neurologist/neurophysiologist Henri Gastaut (1915–1995) and his colleagues made the discovery that mu desynchronization also occurs when the subject is immobile, but watching the movements of others (*Tip 7.5*). As we shall see further along in this chapter, this phenomenon has acquired considerable attention of late in relation to hypotheses linking autism with "mirror neurons."

Although evoked potentials can be recorded in relation to movement, the term "motor-evoked potentials" (MEP) is most typically used to refer to signals recorded from muscles, via the electromyogram (EMG), rather than from the brain. One can, however, observe movement related activity in cortical oscillations, as illustrated in *Figure 7.6*. Note that in this figure, the mu rhythm is in the beta range, roughly centered on 20 Hz prior to the SC, and at a slightly higher frequency between SC and GO (*Tip 7.6*). Two other noteworthy features of M1 neurophysiology, evident in single-trial traces as well as in the ERPs, are (1) VEPs in response to the visually presented

Tip 7.5

In 1954, Gastaut not only published this result in the leading scientific journal *Electroencephalography and Clinical Neurophysiology*, but also in a different kind of scholarly journal: the *Revue Internationale de Filmologie*. This journal examined the sociology of cinema, and one can imagine the appeal of this paper, by Cohen-Séat, Gastaut, Faure, and Heuyer (1954), during those heady days of the French "nouvelle vague" of cinema.

Tip 7.6

Additionally, in Parkinson's disease, entrainment of "mu" at 20 Hz has been shown to improve movement kinetics. These are examples of how human-imposed verbal labels (in this case, "the mu rhythm") don't always map cleanly onto the messiness of biology.

spatial cue (*Tip 7.7*); and (2) a “movement-related potential” (MRP; note that these authors avoided the term “MEP”), the initial positive-going inflection of which systematically preceded the onset of arm movement by approximately 175 ms.

And so, we’ve established from scalp and LFP recordings that motor cortex is maintained in a functionally inactive state that is lifted just prior to the execution of a movement, that sensory signals triggering a movement are registered in this region,

Tip 7.7

Due to the absence of systematic responses to the TC, which in some cases included both visual and auditory components, the authors speculate that “stimulus-induced evoked activity in motor cortex is tightly linked to movement preparation, not merely reflecting sensory inputs” (Kilavik et al., 2010, p. 2347).

FIGURE 7.6 Movement-related dynamics in the LFP. Source: Kilavik, Confais, Ponce-Alvarez, Diesmann, and Riehle, 2010. Reproduced with permission of The American Physiological Society (APS).

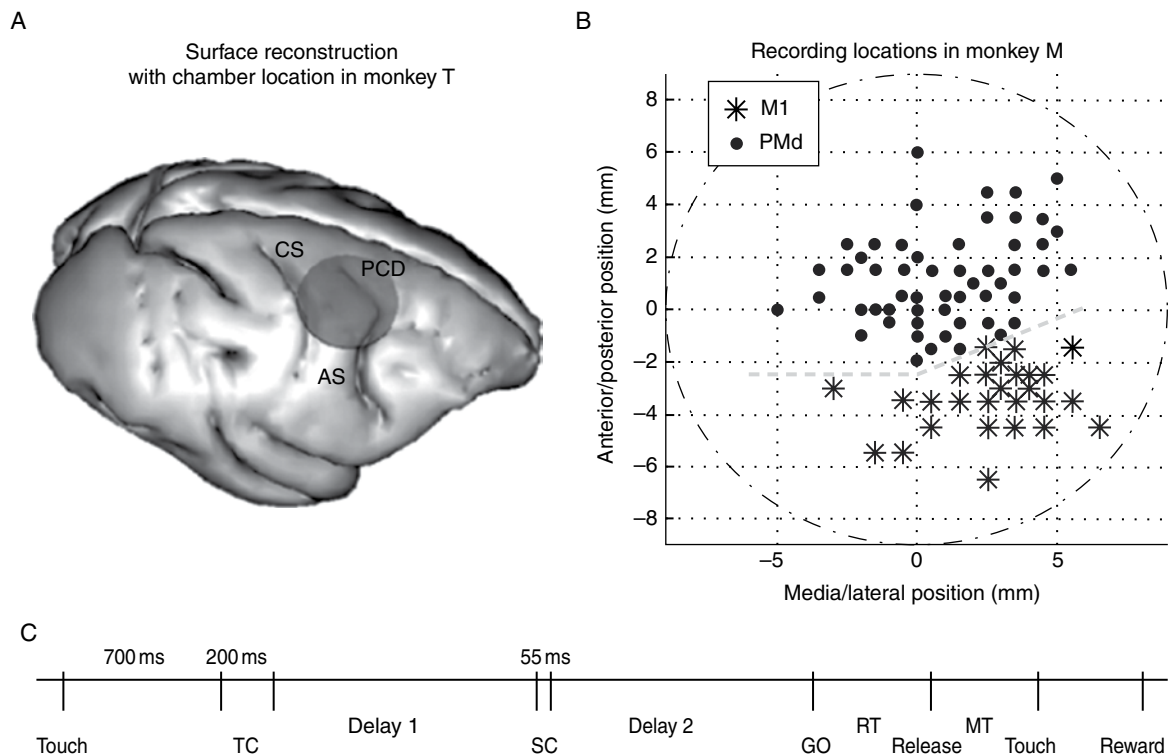


FIGURE 7.6.A Panel **A** is a 3-D reconstruction of a structural MR scan, showing the region from which activity was recorded in one of the two monkeys. CS=central sulcus; PCD=precentral dimple; AS=arcuate sulcus. **B** illustrates recording locations. **C**. Timeline of the task, in which the animals moved a handle (not unlike in *Figure 7.4*) from a central location to one of six radial locations. “Touch” corresponds to when the monkey grasps the handle to initiate the trial; TC is the “temporal cue” that informs the animal whether delay will be short (700 or 1000 ms, depending on the monkey) or long (1500 or 2000 ms); SC=the spatial cue, which indicates the target to which the handle must be moved; GO=the signal to initiate movement; RT=“response time,” defined as the lag between the GO signal and when the handle began to move (i.e., movement-initiation time); MT=movement time, the time between the RT and the acquisition of the target.

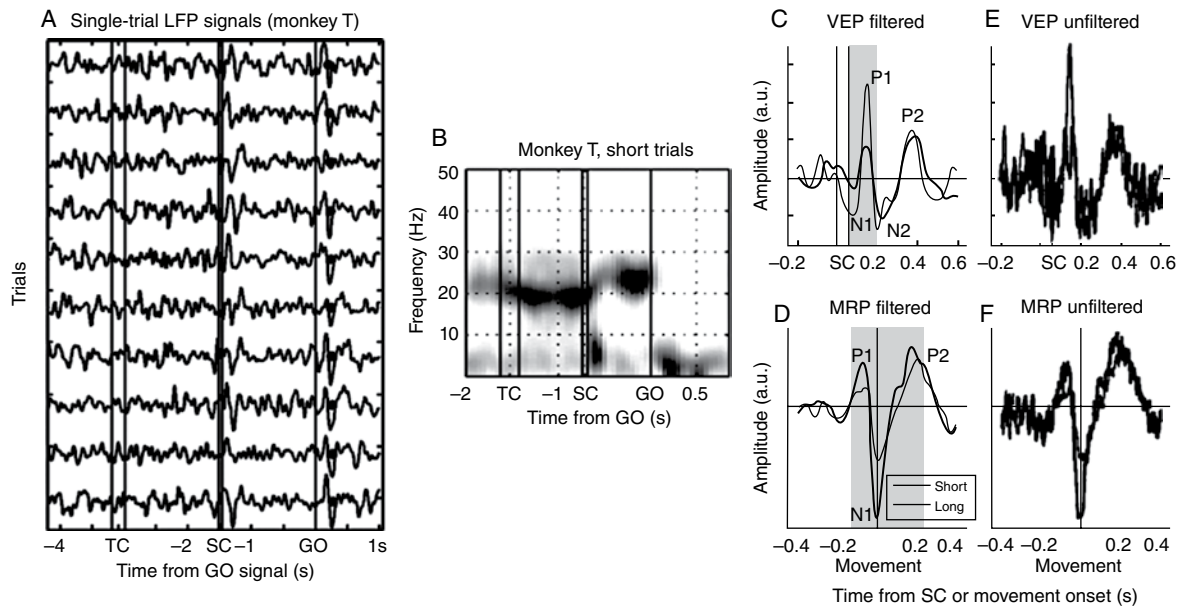


FIGURE 7.6.B Movement-related dynamics in the LFP. **A** illustrates traces from individual trials, bandpass filtered from 1–15 Hz (to emphasize the ERPs), time-locked to the GO signal, and ordered by increasing RT (filled circle following GO signal). **B**. Spectral representation of trials from **A**. **C** and **D**. VEP and movement-related potential (MRP), respectively, time-locked to SC and beginning of MT, respectively, filtered from 1–15 Hz. **E** and **F**. Same data as **C** and **D**, but unfiltered.

and that a large ERP immediately precedes the movement. Next, let’s “drill down” to the level of the individual neuron.

Single unit activity reveals population coding

Among the pyramidal cells of layer V of M1 one finds massive Betz cells, discovered by the Ukrainian anatomist Vladimir Alekseyevich Betz (1834–1894) only a few years after the pioneering microstimulation studies of Fritsch and Hitzig (1870). These are the largest neurons in the nervous system, with large-diameter axons seemingly specialized to send neural impulses very rapidly and over long distances. (Subsequent research has established that many other neurons contribute to the corticospinal tract, with axons from Betz cells contributing only between 3% and 10% of the total.) The breakthrough insight about how these neurons actually generate coordinated movement had to wait another 100 years, for the research of Greek-born, US-based neurophysiologist Apostolos Georgopoulos and his colleagues (*Tip 7.8*).

Tip 7.8

After receiving his MD and PhD at the University of Athens, Georgopoulos moved to Johns Hopkins University in Baltimore, Maryland, to train in the techniques of single-unit extracellular electrophysiology with the pioneering neurophysiologist Vernon Mountcastle. Mountcastle had also mentored Hubel and Wiesel during their time at Johns Hopkins University in the late 1950s/early 1960s.

Georgopoulos, Kalaska, Caminiti, and Massey (1982) recorded from the arm area of M1 from five hemispheres of four different monkeys while they performed a “center-out” task virtually identical to the one described in *Figure 7.6.A*. The results, illustrated in *Figure 7.7*, revealed another important difference between M1 and the primary sensory areas that we have studied to date: instead of narrowly tuned response properties mirroring the precision of receptive fields of, for example,

FIGURE 7.7 The electrophysiology of primary motor cortex.

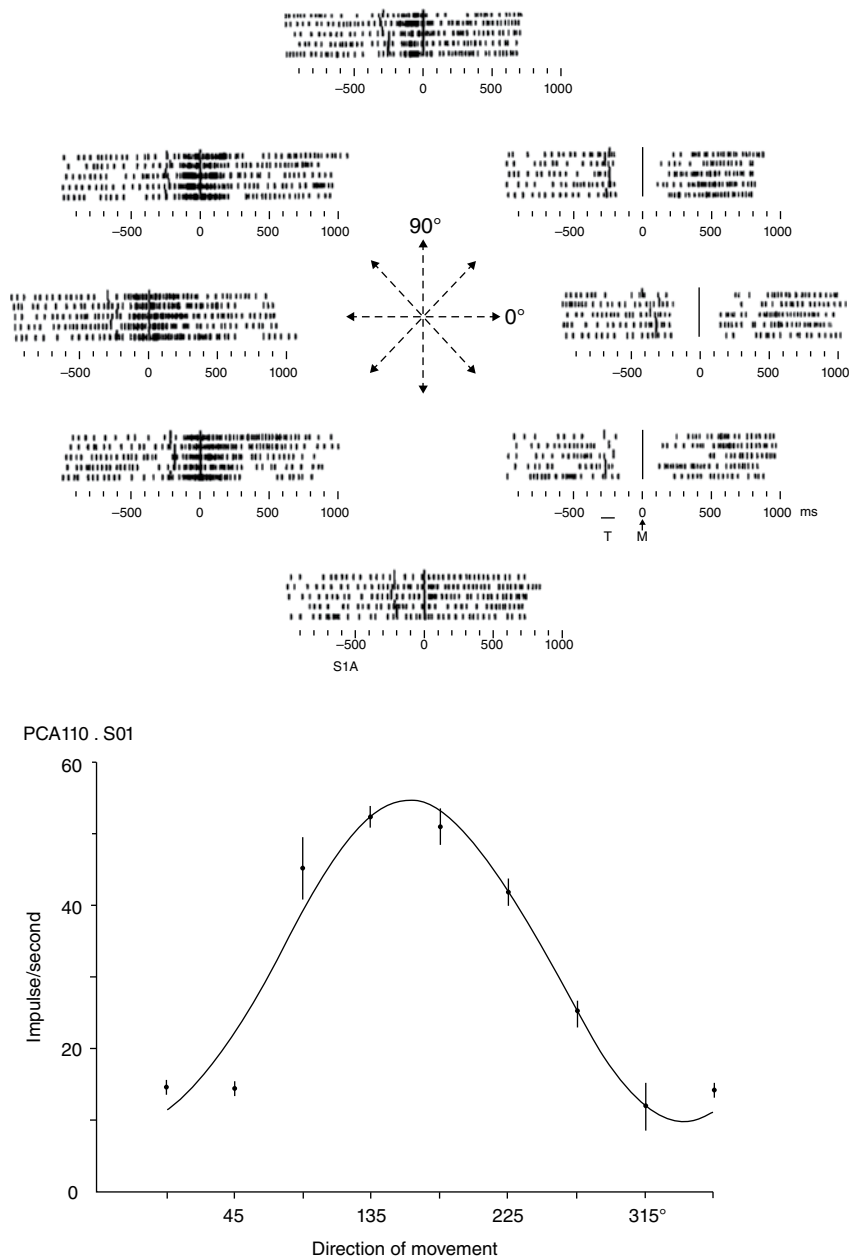


FIGURE 7.7.A Top displays responses of a single neuron from the arm region of M1 for each of five reaches made in each of eight directions, aligned to the time of initiation of movement (arbitrarily labeled “0”). Labels under the raster plot corresponding to movement in the 315° direction indicate the time of the GO signal (labeled “T”) and initiation-of-movement time (M). Below is the tuning curve derived from these data. Source: Georgopoulos, Kalaska, Caminiti, and Massey, 1982. Reproduced with permission of the Society of Neuroscience.

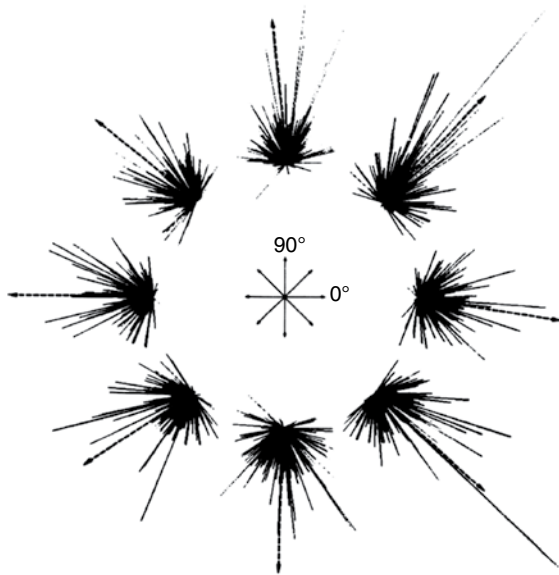


FIGURE 7.7.B Vector representation of the activity of 241 neurons recorded from the arm region of M1, for each of eight directions of movement (center of figure). Dashed lines with arrowheads are vector sums derived from each of the eight directions of movement. Source: Georgopoulos, Caminiti, Kalaska, and Massey, 1983. Reproduced with permission of the authors.

V1 and S1, M1 neurons had remarkably broad tuning. *Figure 7.7.A*, for example, illustrates a neuron whose responses are nearly indistinguishable for movements of 135° , 180° , and 225° (as, indeed, they are for movements of 45° , 0° , and 315°). Different neurons had different overall preferred directions: for example, if *Figure 7.7.A* illustrates a “to-the-left” neuron, others were “away-and-to-the-right” (270° being toward the animal’s body), and so forth. However, all that showed movement-related activity had comparably broad tuning.

Upon first consideration, it’s difficult to see how the activity in neurons with such broad tuning could be responsible for arm movements that we know can be much, much more fine-grained than the separated-by- 45° movements required of the monkeys in this experiment. The principle must be different from what we’ve seen with sensory systems, for which it’s customary to think of individual neurons as acting as feature detectors. Although no one really believes that a single neuron (or column) in V1 can be responsible for a single

visual percept, one can nonetheless imagine, in “short-hand,” that viewing a vertical line preferentially activates neurons tuned for vertical orientations, and that rotating that line ever so slightly to the right will preferentially activate different sets of neurons tuned for a slight tilt to the right. With M1 tuning properties, however, it’d be as though the V1 cell were equally responsive to orientations from 135° to 225° . How could one construct a system with elements whose tuning is so sloppy?! The answer came to the research team after they considered the entire sample of 241 neurons from which they recorded in the arm area in M1: *population coding*.

Instead of thinking of the motor system as being made up of discrete “channels,” each one specialized for its own narrowly defined features, *Figure 7.7.B* illustrates a fundamentally different principle of neural coding. It presents a graphical representation of the response of each of the 241 neurons to each of the eight directions of motion. Each line in each of the eight, radially arrayed “explosion” plots is a vector whose length corresponds to the firing rate of a neuron during motion in that direction. (The neuron illustrated in *Figure 7.7.A*, therefore, is represented in *Figure 7.7.B* by a longish vector at the 135° , 180° , and 225° positions, a somewhat shorter vector at 90° , one shorter still at 270° , and with vectors of length 0 at 45° , 0° , and 315° .) The “vectorial hypothesis” proposed by Georgopoulos, Caminiti, Kalaska, and Massey (1983) makes the assumption that each neuron’s activity on any given trial corresponds to the magnitude of its influence on the direction of movement on that trial. Now recall from mathematics that a vector is a way of representing a force that has a magnitude and a direction, and that if two or more forces are simultaneously present, their vectors can be combined and the resultant “vector sum” (or “vector average,” depending on which operation one performs) will indicate what their combined effects will yield. This is what the dashed lines with arrowheads amid each vectorial plot represent: the empirically derived summation of all 241 vectors. Given that each vector was estimated from only five or so movement trials, and that only 241 neurons were sampled out of the millions that make up motor cortex, it’s astoundingly impressive that an analysis of these neural data could produce an estimated direction of motion (the eight vector sums) that so closely corresponds to the actual direction in which the monkey moved its hand on each trial (the eight arrows in the center of the plot). What Georgopoulos and

colleagues accomplished was to decipher a component of the neural code for skeletomotor control! And the key insight, one to which we will return often for the remainder of this book, is that the principle underlying this function of the nervous system can only be understood by simultaneously considering the activity of a large population of neurons. Serially inspecting one after the next after the next would never reveal it.

I can imagine that to some readers this may seem too abstract. *Vector summation? What, is there some little homuncular mathematician perched up on the precentral gyrus, legs dangling into the central sulcus, calculating vector sums prior to each movement?* Nope. No need in this scheme to resort to the homunculus. Assuming that these are all layer-V pyramidal cells, all located in the arm area, each neuron's action potentials travel to the same general area in the spinal cord, many of them destined for the same motor neuron pool. These action potentials all activate the very same networks of interneurons that were described in the earlier section on *Motor computations in the spinal cord*.

In effect, it's as though 241 teensy weensy micro-electrodes have been simultaneously lowered into the spinal cord. Recall that we saw previously that whereas stimulation of point A in the cord moved the forepaw to location X,Y , and stimulation of point B in the cord moved the forepaw to location X',Y' , simultaneous stimulation of A and B moved the forepaw to X'',Y'' . Thus, the vector summation that a scientist can compute through the arduous analyses and computations summarized here turns out to be a more formal description of what we had glibly summarized earlier: *the motor cortex issues the general command* (in the form of a population code), *and the spinal cord "takes care of the details."*

One of the most exciting consequences of "cracking the code," as Georgopoulos and colleagues have done with motor cortex, is that one can then *decode* new signals as they are generated. *Real-World Applications 7.1* illustrates the application of neural decoding to the exploding field of brain-computer interface (BCI).

REAL-WORLD APPLICATIONS

7.1 The brave new world of BCI

"That's one small step for a robot and one giant leap for a primate," quipped Miguel Nicolelis as he stood in his laboratory at Duke University while a monkey across the room walked on a treadmill. It was January, 2008, and the "primate" that the Brazilian-born neuroscientist was referring to was himself, his collaborators, and, in a sense, the thousands of individuals who may one day benefit from BCI research. All eyes were on a computer monitor in the room, where the "small steps" of a robot that was, at that very moment, walking on a treadmill in Kyoto, Japan, were being displayed via a live video link. What was astounding was that the robot's movements were being controlled by signals from the brain of Nicolelis's monkey. Neural signals generated by a motor cortex located in a Durham, North Carolina, laboratory were being processed and transmitted in real-time via high-speed internet connection to a laboratory in Kyoto, where they were fed to the silicon "motor cortex" of a robot, and thereby making it produce the same movements. Needless to say, the development of methods for brain-computer interface (also referred to as "brain-machine interface," BMI) provides a compelling, concrete demonstration of the relevance of basic research in cognitive neuroscience for many broader societal concerns (*Tip 7.9*).

The example highlighted in *Figure 7.8* is somewhat simpler than the "walking robot" example, and builds on what we've already learned about the motor cortex and how it is studied – instead of moving a handle to a location, it required moving a joystick that controlled the movement of a cursor on a screen. A second twist was that while the monkeys learned to perform these tasks, a computer was analyzing

Real-World Applications 7.1 Continued

task-related neural activity from their frontal and parietal cortices, in order to learn to decode it. Indeed, once their statistics indicated that the computer had “cracked the code,” the researchers disconnected the joystick from the system, such that the movement of the cursor was now being controlled by the monkey’s neural output – “brain control mode.” (Carmena et al. (2003) observed that whereas at first the monkeys continued to move the joystick, “soon they realized that these [movements] were not necessary and ceased to produce them.”)

The final step was to add the robotic arm and “gripper” into the mix by reconnecting the joystick to the computer display, but “giving it” to the robotic arm. Now, to get the cursor to move, the monkey had to control the robotic arm with its “thoughts” (*Tip 7.10*). *Figure 7.8.A* shows that performance on the task dropped precipitously on the first session when “robot control mode” was introduced, but that already by the subsequent session performance improved markedly.

In addition to the impressive engineering feat that this represents, successful decoding necessarily means that one has gained knowledge about how the brain works. Additionally, analyses of neural recordings from this and other experiments have demonstrated how the neural code can change as an animal learns new tasks (*Figure 7.8.B*). This is an example of motor learning, a topic that will be considered further along in this chapter.

FIGURE 7.8 A closed-loop BCI system for direct neural control of a robotic arm. Source: Carmena, Lebedev, Crist, O’Doherty, Santucci, Dimitrov, Patil, Henriquez, and Nicolelis, 2003. Reproduced with permission of the authors.

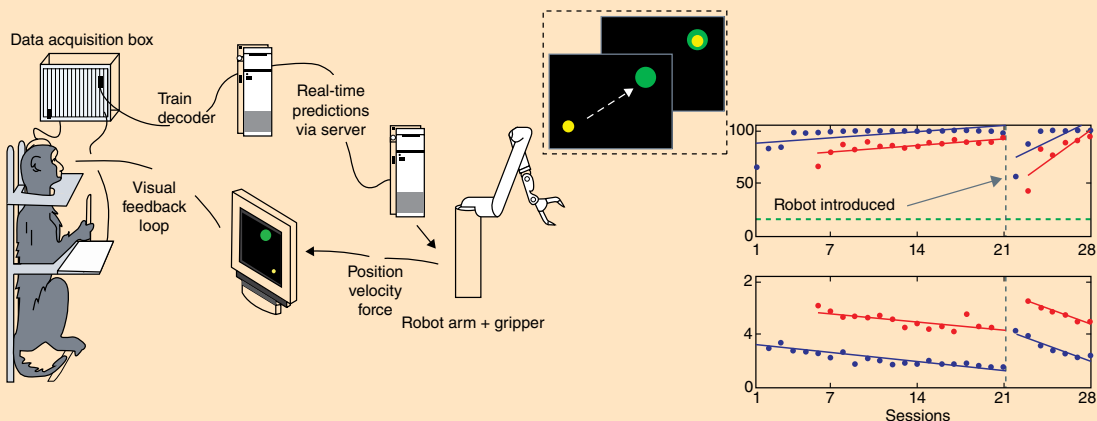


FIGURE 7.8.A The experimental set up is illustrated on the left side of the image, with “Real-time predictions via server” referring to the statistical assessment of the quality of decoding on a trial-by-trial basis (i.e., *How well would we have done on this trial had we guided the cursor with neural signals rather with the joystick?*). The task is illustrated in the dashed box, and the monkey’s performance with joystick control (blue dots and regression line) vs. with neural control (red dots and regression line) as a function of training session. “Robot introduced” indicates the switch to robot control mode.

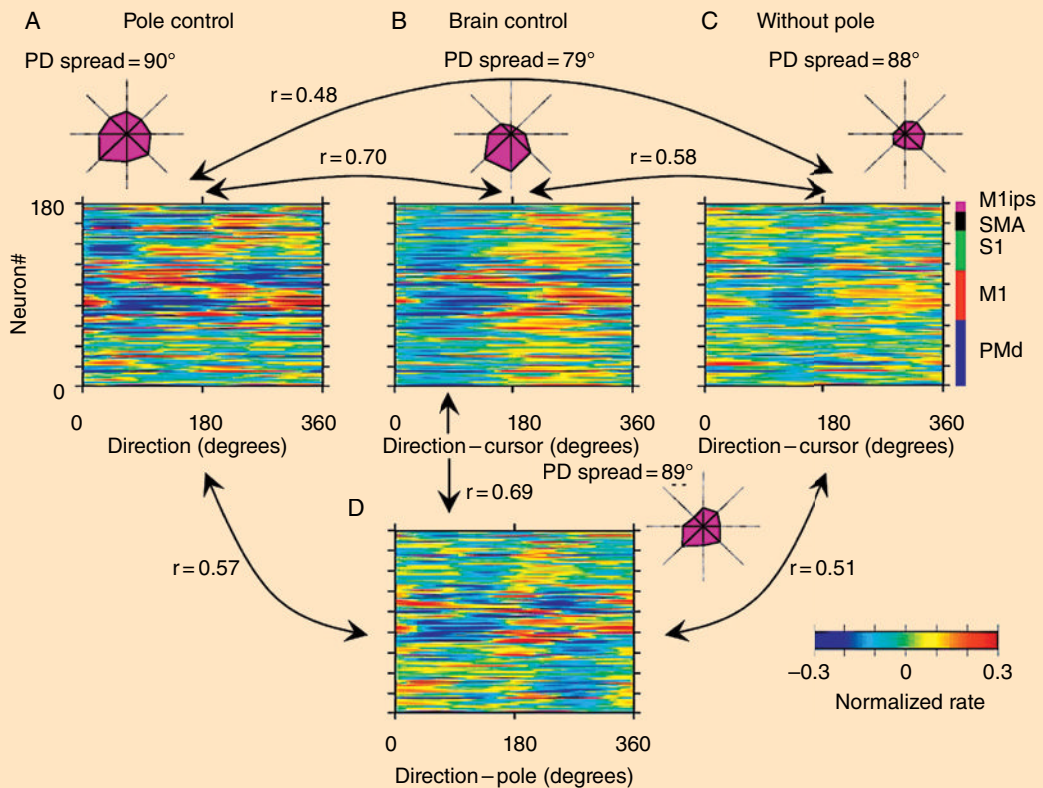
Real-World Applications 7.1 *Continued*

FIGURE 7.8.B Tuning curves for 180 neurons recorded from dorsal premotor cortex (PMd), M1, S1, and SMA (all contralateral to the joystick-controlling hand), and ipsilateral M1 (M1ips). [To understand the graphical convention in these plots, consider the tuning curve in *Figure 7.7.A*. If you were to look down on it from above, you'd want to color code its height at each point along the direction axis, so as to not lose this information. Otherwise it would just look like a straight black line. (The same logic applies for 2-D geographical maps that indicate the elevation of mountains vs. valleys with different colors). Here, red indicates the highest firing rate ("altitude" in the geographical map analogy), violet the lowest. The reason for adopting this "viewed from above" display convention is that it allows the simultaneous display of an entire "stack" of tuning curves in one panel.] The plot on the left shows the tuning curves for this set of 180 neurons when the animal is performing the task in joystick (a.k.a. "pole") control mode, tuning calculated from cursor movements; the plot in the middle from brain control mode but with the joystick still in the hand, tuning calculated from cursor movements; the plot below it is also from brain control mode with the joystick still in the hand, but with tuning calculated from joystick movements; and the plot on the right from brain control mode with no joystick (tuning necessarily computed from cursor movements). Visual inspection suggests that the tuning of many of the neurons in this sample changes as a function of task condition. The r values are coefficients produced by correlating the overall similarity of direction tuning of these 180 neurons in each of four conditions. Finally, the purple polar plots illustrate the tuning of the entire sample in each condition, a "PD spread" of 90° indicating a uniform distribution, and the magnitude of the values at each axis (relative to the origin) the strength of tuning preference in that direction.

Tip 7.9

For example, at the time of this writing, Nicoletis and collaborators based in Germany are working on a robotic exoskeleton that can be worn by a paralyzed patient and controlled by the patient with her “thoughts” (*Tip 7.10*), thereby allowing her walk and interact with her environment in a manner that might at least approximate her abilities pre-injury.

Tip 7.10

The scare quotes highlight the fact that this is imprecise language because, of course, it's neural activity that controls cursors and hypothetical exoskeletons.

MOTOR CONTROL OUTSIDE OF MOTOR CORTEX

Parietal cortex: guiding how we move

One of the facts made clear by studies such as that summarized in *Figure 7.8* is that the neural code involved in carrying out complex actions, such as a reach followed by a grasp followed by a pull, entails much more than the vector summation of 300 or so M1 neurons. (E.g., with just that information, I could be made to move my arm so as to knock the cup of coffee off the arm of my couch, but not to grasp it and bring it to my lips.) Many regions other than M1 can contribute valuable signals for decoding movement-related neural activity. Although the study of Carmena et al. (2003; *Figure 7.8*) didn't venture posterior of S1, others, including Richard Andersen and colleagues (of *Chapter 6* renown), have done so, emphasizing that a greater range of cognitive factors – including intention, expectation, and reward value – might be best sampled outside of traditional motor areas. “For example, PRR [parietal reach region; see *Figure 6.11*] neural activity codes the intention to reach to an object at a particular location in space, whereas motor

cortex codes the direction to move the hand” (Andersen, Musallam, and Pesaran, 2004, p. 271; *Tip 7.11*). The critical role for PPC for the planning and execution of visually guided action, through its contributions to sensorimotor integration and coordinate transformation, also figure prominently in the alternate framing of the ventral vs. dorsal pathways as being one of *what* vs. *how*, as summarized in *Research Spotlight 7.1*.

Cerebellum: motor learning, balance, ... and mental representation?

From the standpoint of aesthetics, the cerebellum may be the most beautiful part of the brain: the tight foliation of its cortex (housing 10% of the volume but > 50% of the neurons of the brain); the remarkably profuse, yet “flat,” dendritic arbors of its Purkinje cells; the elegant simplicity of its circuitry (*Tip 7.12*). As summarized previously, the cerebellum is a hub in one of the two major recurrent circuits of the motor system, serving an online error-monitoring function and feeding real-time performance information back onto frontal cortex.

Tip 7.11

The review by Andersen, Musallam, and Pesaran (2004) also emphasizes that LFPs can be just as valuable a signal for BCI as is spiking activity. Indeed, many applications of BCI also use such signals as the mu rhythm, and some have been successful using extracranial EEG alone.

Tip 7.12

The principles of cerebellar function have both inspired, and been inspired by, engineering control theory and robotics. Indeed, a seminal book from 1967 that influenced almost all research to follow, co-authored by Nobel-laureate neuroscientist Sir John Eccles, was entitled *The Cerebellum as Neuronal Machine*.

RESEARCH SPOTLIGHT

7.1 A neurological dissociation between perceiving objects and acting on them

A patient with the initials D. F. developed a profound impairment in the ability to recognize objects (a form of visual agnosia, a syndrome that we will define and consider in detail in *Chapter 9*) after carbon monoxide inhalation produced extensive damage in lateral occipital BAs 18 and 19, but largely sparing V1. Despite this severe impairment, she seemed remarkably unimpaired when “directing accurate reaching movements toward objects in different orientations” (Goodale et al., 1991, p. 155). To formalize and quantify this dissociation, which was reminiscent of that demonstrated by monkeys with IT lesions in the Pohl (1973) study (*Methodology Box 5.1*), Goodale and colleagues (1991) contrasted her ability to visually perceive the orientation of a slot in an upright disk versus her ability to “post” a card through the slot (i.e., to slide it through as though putting a postcard into a mail slot).

Panel A in *Figure 7.9* illustrates the responses on five trials at each of six different orientations in a “perception” task, in which the patient and two control subjects were asked to match the orientation of a slot in a disk that was 45 cm distant. At all angles except 90° (i.e., vertical) D. F. was profoundly impaired. (And because each trial was started with the card held at 90°, her relatively intact performance at this orientation

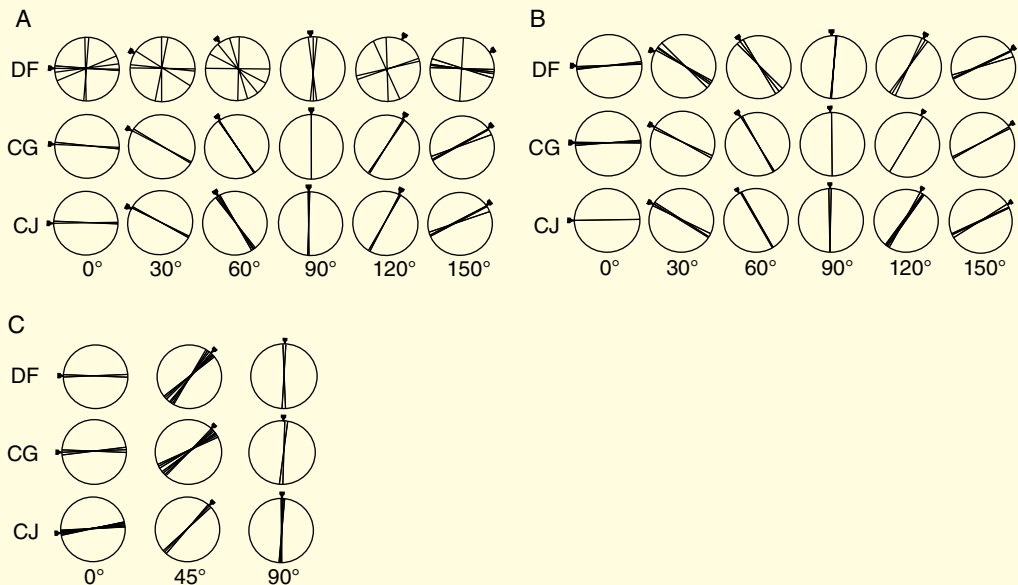


FIGURE 7.9 Perception vs. action in patient D. F. **A** illustrates the responses on five trials at each of six different orientations in which the patient and two control subjects (C. G. and C. J.) were asked to hold a card at the same orientation as the slot, whose true orientation is marked on each disc with a small arrow-head. **B** illustrates the orientation of the card just before it made contact with the disc. **C** illustrates an eyes-closed version of **A**. Source: Goodale, Milner, Jakobson, and Carey, 1991. Reproduced with permission of Nature Publishing Group.

Research Spotlight 7.1 *Continued*

may be artifactual.) **B** illustrates that, in contrast, when asked to “post” the card by reaching to insert it into the slot, her performance was nearly as precise as that of the control subjects. **C** illustrates a mental imagery version of the matching task, in which subjects were asked to close their eyes and imagine the slot at each of three orientations. This indicates that poor performance in **A** was not due to a misunderstanding of instructions, or some other uninteresting factor.

A second experiment looked at a second kind of visually guided action, grasping. First subjects were asked to indicate with finger and thumb the front-to-back size of each of five plaques, ranging in shape from a 5 x 5 cm square to a 2.5 x 10 cm rectangle. These two digits were fitted with infrared light-emitting diodes so that precise measurements could be made of finger-to-thumb aperture. The results showed that, unlike control subjects, D. F.’s responses did not differ with shape. When asked to pick each of the shapes up, in contrast, she was just as accurate as control subjects.

In interpreting their findings, the authors conjectured that they did “not correspond to the two streams of output from primary visual cortex generally identified in the primate brain – a ventral one for object identification and a dorsal one for spatial localization [... instead, they] indicate separate processing systems not for different subsets of visual information, but for the different uses to which vision can be put.” Therefore, to return to a theme from the previous chapter, this view has more in common with *intention* rather than *attention* interpretations of the neglect syndrome and of the functions of the parietal lobe.

(See *Further Reading* for references to lively debate about this idea that continues more than two decades after its initial articulation.)

Cerebellar circuitry

The cerebellum is highly modular, consisting essentially of a bank of hundreds of thousands of parallel circuits that all perform the same computational operations, but do so on different, segregated channels of information. The central processing element around which cerebellar circuits are built is the Purkinje cell. These are among the largest cells in the nervous system, with massive dendritic arbors that afford an incredibly large number of synaptic connections (estimated to be as high as 200,000 per neuron!). Despite their size and extensive and intricate branching, however, these arbors only extend through a 2-D plane. A row of Purkinje cells aligned in a “strip” that is parallel to the plane of their dendritic arbors makes up a circuit (called a “micro-zone”), the output of which, via Purkinje cell axons, is focused on neurons of the cerebellar deep nuclei that were discussed in the introductory section of this chapter. At a simplified level, the copy of the efferent motor command is carried into the

cerebellum by climbing fibers, and information about the sensory inputs – the “context” in which an action is taken and the sensory consequences of that action – by mossy fibers. Climbing fibers are axons originating in the inferior olive in the brainstem, their name deriving from the fact that a single climbing fiber wraps around and around the basal dendrites of a single Purkinje cell, like a vine climbing a tree trunk. Climbing fibers make between 300–500 synaptic connections with a single Purkinje cell. Mossy fibers arise from the pons (in the brainstem), their name deriving from the distinctive “rosettes” in which they form synaptic connections with granule cells. Granule cells then send “parallel fibers” that run perpendicular to, and therefore through, the dendritic arbors of many Purkinje cells (i.e., they cut across many microzones). Individual parallel-fiber synapses on Purkinje cells are relatively weak, whereas a single action potential in a climbing fiber will reliably trigger an action potential in a Purkinje cell.

Cerebellar functions

An influential model of the computations carried out by the cerebellum, articulated by the British neuroscience polymath David Marr (1945–1980), and subsequently refined by the American electrical engineer and roboticist James Albus (1935–2011), is summarized as follows. The premise, rooted in control theory from engineering, is that the Purkinje cells in the cerebellum contain a representation of a prediction of what the effects of a movement should be (i.e., an “internal model”). The sensory report of what the consequences of the action actually were is delivered, via parallel fibers, to the cerebellum. To the extent that the internal model and the actual consequences differ, a modification is made to the cerebellar circuitry, and the results of this assessment get forwarded to the motor cortex so that it can adjust how it commands this action the next time. In effect, the idea is that the cerebellum houses one instantiation of the predictive coding models introduced back in *Chapter 3*. In his influential 1971 paper, Albus summarized the role for the modification of cerebellar circuits in motor learning as follows:

It is an obvious fact that continued training in motor skills improves performance. Extended practice improves dexterity and the ability to make fine discriminations and subtle movements. This fact strongly indicates that learning has no appreciable tendency to saturate with overlearning. Rather, learning appears to asymptotically approach some ideal value. This asymptotic property of learning implies that the amount of change that takes place in the nervous system is proportional to the difference between actual performance and desired performance. A difference function in turn implies error correction, which requires a decrease in excitation upon conditions of incorrect firings. (Albus, 1971, p. 52)

By the Marr–Albus model, “desired performance” is represented by the climbing fibers, which carry the copy of the efferent motor signal. When discharge of a climbing fiber triggers an action potential in a Purkinje cell, there is a refractory period that follows, during which the Purkinje cell cannot fire another action potential. If action potentials from parallel fibers arrive during his refractory period, the consequence of this is a weakening of synapses between them and the Purkinje cell (*Tip 7.13*). This modification of synaptic strength in

Tip 7.13

We haven’t yet, nor will we for another few chapters, introduced principles of synaptic plasticity (i.e., the changing of the efficacy of synapses) that underlie learning and memory. For now, suffice it to say that the converse of the aphorism that “neurons that fire together wire together” is that connections between neurons whose activity is decorrelated tend to be weakened.

the cerebellum is a neural instantiation of motor learning. One way to think of this process is that subtle changes in cerebellar circuitry might produce subtle changes to the length of a few of the M1 neuronal activity vectors illustrated in *Figure 7.7.B*, which, in turn, would produce an ever so subtle change in the endpoint of that movement the next time it is executed.

The circuitry of the cerebellum continues to fine-tune our movements throughout our lives. Evidence for this comes from the fact that damage to the cerebellum produces abnormalities of posture, gait, reaching, and other skeletomotor functions regardless of when in life the damage is incurred.

Interestingly, some channels of the cortico–cerebellar circuit link PFC with the cerebellum, and so this computational principle presumably also carries out fine-grained, online corrective processing of high-level cognitive functions, such as those mentioned in the introductory paragraph to this chapter. The Japanese neurophysiologist Masao Ito who, as a younger man, co-authored the book highlighted in *Tip 7.12*, has proposed an intriguing model that we can’t detail here, but whereby, in addition to its internal models of motor behaviors, the cerebellum may also encode internal models “that reproduce the essential properties of mental representations in the cerebral cortex ... [thereby providing a] mechanism by which intuition and implicit thought might function” (Ito, 2008, p. 304).

Basal ganglia

The basal ganglia define the second major recurrent circuit of the motor system. This circuit, too, is not limited to motor control. Rather, as *Figure 7.3.A*

highlights, the skeletomotor circuit is just one of five frontal cortical-basal ganglia circuits. One major difference between the skeletomotor circuit and the other four is that its cortical efferents synapse in the putamen of the neostriatum, whereas three of the others target different regions of the caudate nucleus of the neostriatum, and the ACC and MOFC of the limbic circuit target the nucleus accumbens of the neostriatum (also referred to as the ventral striatum). Subsequently, each of the five circuits, although remaining segregated, bifurcates into a direct and an indirect pathway, the direct linking striatum to the internal segment of the globus pallidus (GPi), the GPi to thalamus, and thalamus back to cortex.

A subcortical gate on cortical activity

Although learning the wiring diagram of this circuit has been known to induce dizziness and headache in some students, the basic operating principles are actually quite straightforward. The key to understanding its function is remembering a distinctive property of the GPi. The GPi is a tonically *active* structure. That is, its baseline state, when the circuit is not being engaged, is to steadily fire at a high frequency (~85 Hz in one study in the monkey; this will vary with species, and with an animal's overall physiological state). Because GPi neurons are GABAergic, the default state of affairs is that the GPi maintains a steady inhibitory “clamp” on the thalamus. Now, moving one step upstream, the cells of the neostriatum that project to the GPi are also GABAergic. Thus, engaging the direct pathway in a basal ganglia circuit has the effect of phasically inhibiting the GPi's activity, thereby phasically lifting the GPi's inhibitory clamp on a small region of thalamus, and thereby allowing it to send an excitatory signal to its frontal cortical target (*Tip 7.14*). In this way, the basal ganglia can be construed as “gating” cortical activity. (We'll see in subsequent chapters that this mechanism has been proposed to be of central importance to allowing us to, for example, change behavioral strategies from one to another.)

The role of dopamine

Figure 7.2 illustrates another important component to the operation of the basal ganglia: the delivery of the neurotransmitter dopamine (DA). DA, along with norepinephrine, serotonin, and others, is classified as

Tip 7.14

The basal ganglia's relationship with the thalamus is reminiscent of the feedback from primary sensory cortex onto sensory relay nuclei of the thalamus (e.g., LGN, MGN, and VP); note, however, that the basal ganglia only synapse on “higher-order” nuclei of the thalamus that project to frontal cortex, including VL, VA, and MD.

Tip 7.15

In *Chapter 14* and *Chapter 15*, however, we'll see that midbrain DA systems are intimately involved in cognitive flexibility and in the signaling of the rewarding status of a stimulus or a behavior.

a neuromodulatory neurotransmitter. That is, it's typically not thought to convey information in the way that glutamatergic neurons do (*Tip 7.15*), but it sets the “tone” within networks, thereby biasing the way in which glutamatergic and GABAergic signals are processed. DA is synthesized in three adjacent nuclei in the midbrain. Depending on who is describing them, they are either identified as *A8*, *A9* and *A10*, or as the substantia nigra pars compacta (SNpc) and the ventral tegmental area (VTA). The SNpc (or A9) delivers DA to other nuclei of the basal ganglia, most densely to the neostriatum via the **nigrostriatal tract**, the other nuclei projecting primarily to neocortex.

The influence of DA on the basal ganglia can be considered at many levels, and here we'll begin with what's illustrated in *Figure 7.2*. Although all fibers of the nigrostriatal tract deliver “the same” molecule throughout the neostriatum, its effects can be either excitatory or inhibitory, depending on the zone receiving this input. How can this be? The answer is that, unlike glutamate and GABA, DA receptors are not linked to membrane-spanning channels that directly gate the passage of ions. Rather, they are so-called “second messenger-linked” receptors. What this means is that the binding of a molecule of DA to a DA receptor initiates a cascade of chemical

events within the postsynaptic neuron, the ultimate consequences of which might be excitation or inhibition, depending on which second-messenger pathway is triggered. There are two classes of DA receptor: D1-related and D2-related. In general, D1-related receptors produce a depolarizing effect, D2-related receptors a hyperpolarizing effect.

The neostriatum is made up of a checkerboard of compartmentalized zones that are dominated by D1-related or by D2-related receptors. The two types of zones are the basis for the direct vs. the indirect pathways, because their outputs project to either the GPi or the external segment of the globus pallidus (GPe), respectively. One of the ways that we understand the influence of DA on this system is from decades of research on Parkinson's disease (PD), which is characterized by the progressive degeneration of the SNpc, and, therefore, a progressive lowering of DA levels in the neostriatum. Motorically, PD is associated with three cardinal clinical symptoms: resting tremor in one or more limbs, muscular rigidity, and

slowness to initiate movement ("bradykinesia"; see *Web Link 7.1* for video demonstrations of patients with PD). One way to understand these symptoms (and, by extension, to understand the role of DA in the healthy system) is illustrated in *Figure 7.2*: the reduction of DA in the neostriatum throws off the balance of its output, such that the descending signal of the direct pathway is weakened, and that of the indirect pathway is strengthened. The direct pathway is consequently less able to impose phasic inhibition on the GPi, which, in turn, reduces excitatory drive from thalamus to cortex. This can account for the rigidity and bradykinesia seen in PD. By applying the same analysis to the GPe one can see that a similar endpoint is reached. (*Web Link 7.2* considers Huntington's disease, characterized by progressive degenerative of the caudate nucleus.)

A second level at which DA influences the basal ganglia is in controlling the synchrony between its nuclei. As is detailed in *Real-World Applications 7.2*, there are both pharmacological and surgical

REAL-WORLD APPLICATIONS

7.2 Pharmacological and surgical treatment of PD

For decades, neuropathologists have known that PD is associated with a degeneration of the SNpc, because postmortem brains of PD patients have visibly less black-pigmentation in the midbrain than do brains from non-neurological autopsies. Although dopamine replacement is an obvious idea for treatment, development of an effective drug was held up for years by the fact that DA cannot cross the **blood-brain barrier**, the layer of endothelial cells that prevents the diffusion from capillaries of many substances carried in the bloodstream. A breakthrough occurred when Swedish physician and neuropharmacologist, Arvid Carlsson, who would be awarded the Nobel Prize in 2000, discovered that a precursor to dopamine, L-3,4-dihydroxyphenylalanine (commonly referred to as levodopa, or L-dopa), does cross the blood-brain barrier. Once in the brain, enzymes present in dopaminergic neurons convert L-dopa to dopamine. Shortly after the initial discovery it was shown that L-dopa can be effective at alleviating motor symptoms of PD. To this day, L-dopa remains the primary pharmacological treatment for PD.

However, L-dopa has limitations. One is a result of the fact that the loss of DA neurons does not occur evenly across dopaminergic neurons of the midbrain. Rather, neurons of the SNpc precede those of the VTA, meaning that optimal treatment of motor symptoms with L-dopa (*Figure 7.10*) can result in an excess of DA in other brain regions, such as the PFC, which is innervated by the VTA. Dopamine's effect on the

Real-World Applications 7.2 Continued

brain, like that of many neurochemicals, follows an “inverted U-shaped” function, meaning that neural function is suboptimal when the concentration is too low, optimal in a mid-range, but impaired again as the concentration gets too high. A complication with titrating the dose of L-dopa for PD patients, therefore, is finding a “happy medium” between a dose that provides optimal relief from motor symptoms and one that won’t produce cognitive impairment. A second limitation is that L-dopa can only be effective so long as there is a sufficient number of dopaminergic neurons to take it up and convert it into DA. At advanced stages of progression of the disease, there simply aren’t enough viable neurons left in the SNpc to convert L-dopa into DA. These are reasons why surgery is often needed.

The first surgical treatments developed were “-otomies”; for example, a *pallidotomy* entails lesioning the GPi so as to reduce its pathologically strong inhibitory clamp on the thalamus. A *thalamotomy* most often targets the ventral intermediate nucleus of the thalamus, which has the specific effect of reducing tremor. At present, however, the more common surgical treatment for PD is so-called deep brain stimulation (DBS), in which one or more stimulating electrodes are implanted in a basal ganglia nucleus (often the STN, but, as *Figure 7.10* illustrates, other nuclei can also be targeted). The electrode is driven by a battery that is surgically implanted in the chest, and can be controlled via a handheld magnet. As *Figure 7.10* makes clear, brain function is exquisitely sensitive to the precise oscillatory dynamics within and between basal ganglia nuclei. Steady, high-frequency electrical stimulation delivered to one or more of these can be an effective counter to disease-related loss of DA.

interventions for the treatment of PD, and during surgery it’s possible to record neural activity in the patient. In one study (*Figure 7.10*) the surgical and research team compared recordings when dopamine-replacement medication was withheld from patients (i.e., they were “OFF” meds) vs. when the patients had been administered the optimal level of this medication (“ON”). Recordings from the OFF condition showed that the power in the LFP recorded from the subthalamic nucleus (STN) and GPi was shifted to lower frequencies relative to the ON condition. Even more strikingly, oscillatory coherence of LFPs between the two structures was highest at 6 Hz and 20 Hz in the OFF condition, but shifted to 70 Hz in the ON condition. It may be, therefore, that the resting tremor in PD results from a disruption of optimal oscillatory synchrony in this system. Note, also, that these indices of the oscillatory dynamics within and between structures give a concrete example of what is meant by references to a region’s “dopaminergic tone.”

FIGURE 7.10 The effects of dopamine-replacement therapy on oscillatory dynamics within and between basal ganglia nuclei. Source: Brown, Oliviero, Mazzone, Insola, Tonali, and Di Lazzaro, 2001. Reproduced with permission of the Society of Neuroscience.



FIGURE 7.10.A Axial MR image of a patient after bilateral implantation of electrodes in the GPi (black arrows) and the STN (white arrows).

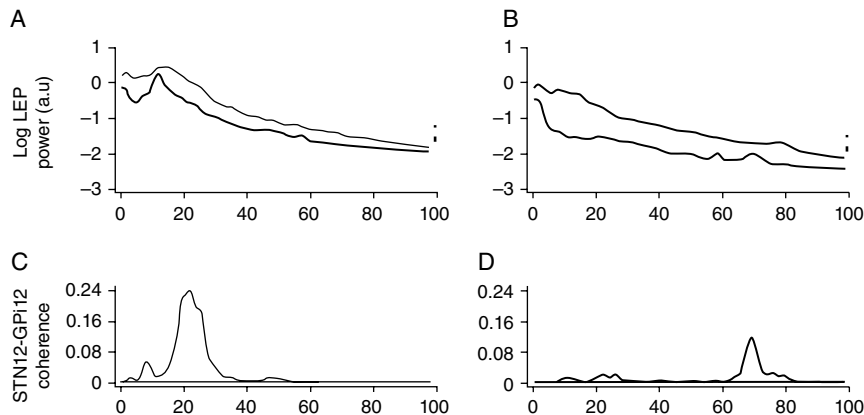


FIGURE 7.10.B Panels **A** and **C** illustrate measurements averaged over four patients when they were OFF L-dopa and **B** and **D** when they were ON. **A** and **B** show power spectra in the GPi (higher power) and the STN; **C** and **D** coherence between the two structures (i.e., the degree to which oscillations in the two structures are synchronized, at each frequency from 0–100 Hz). Brown et al., 2001.

COGNITIVE FUNCTIONS OF THE MOTOR SYSTEM

As suggested at the beginning of this chapter, many have argued that high-level cognition may have evolved over millennia as increasingly abstracted versions of motor control. Ito's (2008) proposal of mental models represented in the cerebellum would fall into this category. Also consistent with this idea is the fact that the cortico-basal ganglia-thalamic circuits of the PFC and the ACC/LOFC operate on the same principles as does the motor circuit that we focused on in this chapter. Two additional examples of this idea have come from the Italian neurophysiologist Giacomo Rizzolatti and his colleagues. One is the "premotor theory of attention," the idea that attention may "have evolved from"/"be a byproduct of" motor preparation. This theory will receive considerable attention in the next chapter, when we take up oculomotor control and the control of attention. In the meantime, this chapter

concludes, via *Research Spotlight 7.2* and *Thought Box 7.1*, with another of Rizzolatti's discoveries: mirror neurons.

IT'S ALL ABOUT ACTION

The principles of motor control lie at the core of understanding human behavior. On one level, this must literally be true, because we cannot *behave* without the release of ACh at the neuromuscular junction causing the contraction of muscle, nor all the steps that precede this event. At a deeper level, we have also seen that circuits identical to those responsible for the triggering of a motor command and for the fine-tuning of its execution apply those same computations to signals from prefrontal areas associated with our most abstract thoughts and our most powerful feelings. It will serve us well, therefore, as we delve further from the sensory and motor peripheries, in exploration of "higher levels" of cognition, to be at the ready with such principles and concepts as population coding, efference copy, and neuromodulatory tone.

RESEARCH SPOTLIGHT

7.2 Mirror neurons (Gallese, Fadiga, Fogassi, and Rizzolatti, 1996)

Sector F5 is a subregion of premotor cortex (*Figure 7.11.A*) that had been shown in previous recording and microstimulation experiments to be involved in the high-level representation of actions involving the hand and the mouth. Different neurons, for example, responded selectively to grasping, tearing, or holding movements. In the study featured here, Gallese and colleagues, upon isolating a neuron, would present the monkey with an item of food on a tray, either within or outside of the animal’s reach, as well as with objects of various sizes. The size manipulation enabled distinguishing between three types of grip: “precision grip” (thumb and index finger); “prehension grip” (thumb opposite fingers to extract an object from a deep, narrow container); and “whole-hand prehension.” This procedure can be construed as analogous to determining the visual response properties of a neuron in IT cortex by presenting it with a series of objects.

The novel innovation of this study was that the testing battery also included an experimenter performing several actions while the monkey watched: grasping a piece of food from a tray, bringing the food to his/her mouth, giving the food to a second experimenter, etc. Of 532 neurons recorded from F5, 92 discharged to specific actions by the monkey *and* to the observed actions of an experimenter. These were dubbed mirror neurons. The actions that mirror neurons responded to were categorized as shown in the table.

Observed hand actions	Number of neurons
Neurons with one preferred action	
Grasping	30
Placing	7
Manipulating	7
Hands interaction	5
Holding	2
Neurons with two preferred actions	
Grasping/placing	20
Grasping/manipulating	3
Grasping/hands interaction	3
Grasping/holding	5
Grasping/grasping with the mouth	3
Placing/holding	1
Hands interaction/holding	1
Neurons with three preferred actions	
Grasping/placing/manipulating	1
Grasping/placing/holding	4

Research Spotlight 7.2 Continued

These responses were insensitive to manipulation of factors related to the animal's reference frame, such as side of the body the actions were performed on, or distance of observed actor. Importantly, however, these neurons did not respond when their preferred action was performed as mimicry (e.g., experimenter pretending to pick up a piece of food when there was no object on the tray; *Figure 7.11.B*), or when observing the same action performed with tools (forceps or pliers) instead of the hand.

In this study, 31.5% of mirror neurons were classified as being “strictly congruent,” meaning that they responded preferentially to the same action, regardless of whether the animal was observing an experimenter performing the action, or executing the action itself. 61% were classified as “broadly congruent” – preferred observed and executed actions were similar, but not identical. For example, one such neuron responded vigorously during observation of a precision grip and observation of whole-hand prehension, but only responded to execution of a precision grip. The remainder were “non-congruent” with regard to preferred observed vs. executed action(s).

The authors suggested that mirror neurons are sensitive to the meaning of an action, rather than to its precise sensory or motoric properties. They further speculated that mirror neurons may underlie learning through imitation (prevalent in humans as well as in nonhuman primates), as well as, possibly, speech perception in humans. This final idea will get some elaboration in *Thought Box 7.1*, and again in *Chapter 18*.

FIGURE 7.11 Mirror neurons in premotor cortex of the monkey.

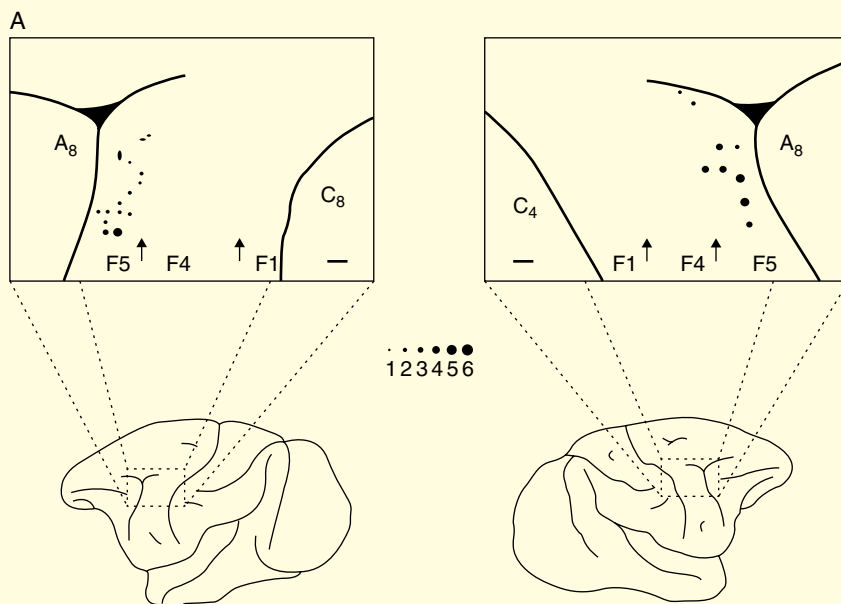


FIGURE 7.11.A Location of electrode penetrations in sector F5 of BA 6. Source: Gallese et al., 1996. Reproduced with permission of Oxford University Press.

Research Spotlight 7.2 *Continued*

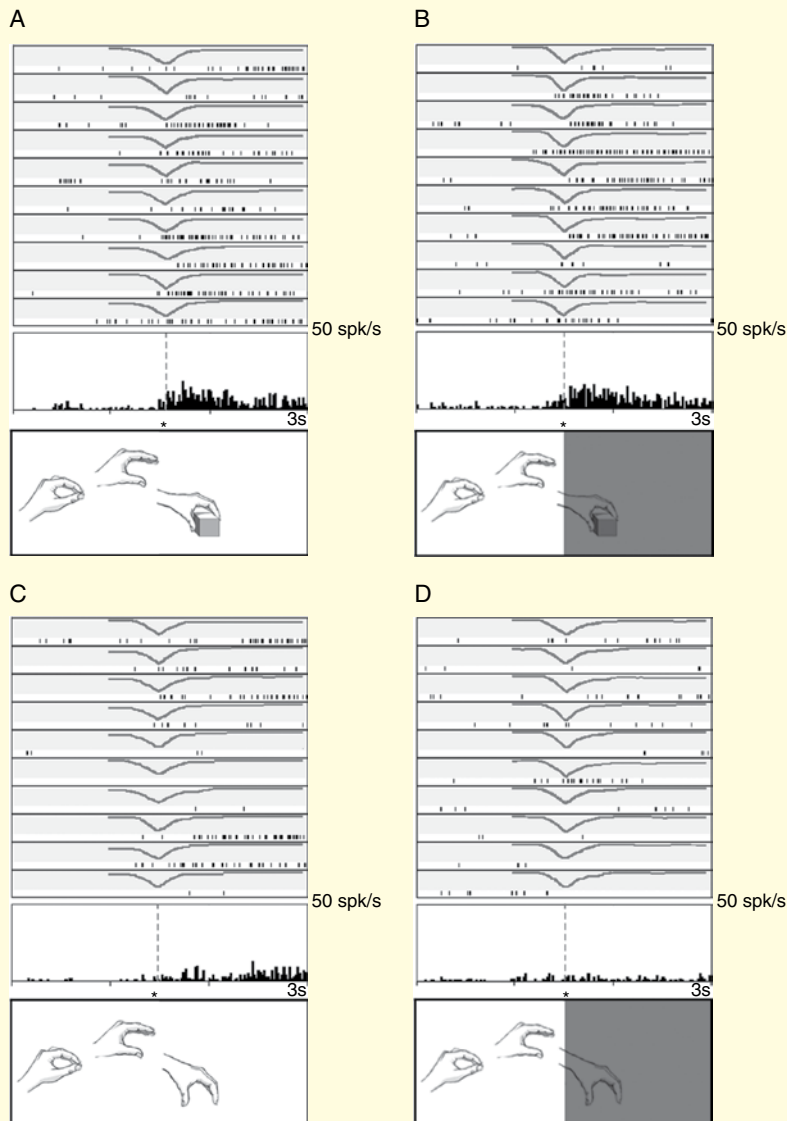


FIGURE 7.11.B Responses of an F5 neuron in a follow-up study in which the monkey watched while an experimenter reached to either grasp an object (panels **A** and **B**) or to mimic the same action but in the absence of the object. In between the monkey and the table on which the object was displayed was a metal frame that could either be empty (panels **A** and **C**) or hold an opaque screen, thereby blocking the monkey's view of this location (panels **B** and **D**). For all four trial types, neural activity is aligned to the moment in time when the reaching hand passed the boundary of the frame (and, therefore, became no longer visible in **B** and **D**). (The waveforms above each row of rasters illustrate the location of the hand relative to the boundary of the frame as a function of time, aligned the same as neural data.) Note how the neuron only responds to the act of grasping a physical object, not to the identical movement when the object was absent, and that it does so regardless of the presence or absence of the screen. Source: Umiltà, 2000. Reproduced with permission.

THOUGHT BOX

7.1 The mirror system: responsible for autism spectrum disorder, speech perception, and the birth of human culture and civilization; or a cognitive neuroscientific canard?

Mirror neurons are among the “hottest” topics of research in contemporary cognitive neuroscience. One factor driving this interest is that many believe that they offer a neural basis for many popular schools of thought in cognitive psychology, including embodied cognition, common coding theory, and the motor theory of speech perception. The latter was referenced by Gallese and colleagues in their paper featured in *Research Spotlight 7.2*. It is a theory, first proposed in the 1960s, that an important component of speech perception is the “listener” watching the mouth of the talker, and simulating in her mind the larynx, tongue, and lip movements being made by the talker. Gallese and colleagues (1996) speculated that the F5 region in the monkey may bear homologies to Broca’s area in the human – the region of left inferior PFC that we first encountered in *Chapter 1*, and will do so again in *Chapter 18*.

With regard to autism spectrum disorder (ASD), some have proposed that dysfunction of the mirror system underlies the disorder. One high-profile study (Dapretto et al., 2005) compared fMRI activity in typically developing adolescents vs. “high-functioning” children with ASD on tasks requiring passive observation vs. mimicking of visually presented facial expressions. The findings indicated that, whereas behavioral performance was comparable between the two groups, “the ASD group showed no activity in the mirror area in the pars opercularis” (Dapretto et al., 2005, p. 29; see my rant in *Chapter 6* about why the assertion of “no activity” cannot, literally, be true). Indeed, despite this statement, the authors further report that the level of activity in this region that purportedly showed “no activity” correlated negatively with ASD subjects’ scores on two scales, indicating that “the greater the activity in this critical component of the [mirror neuron system] during imitation, the higher a child’s level of functioning in the social domain” (pp. 29–30).

Finally, at *Web Link 7.3*, you can hear V. S. Ramachandran (whose research we first encountered in *Chapter 4*, in relation to phantom limb and referred sensations) articulate his idea that mirror neurons are at the root of the explosion of human culture and civilization that began roughly 100,000 years ago. “Even though the human brain had achieved its current size roughly 3–400,000 year ago,” he notes, “something [neurologically] important must have happened” 100,000 years ago to trigger the rapid development and propagation of tool use, fire, shelter, language, and so on. In his talk, Ramachandran argues that that “something” was “the sudden emergence of a sophisticated mirror neuron system, which allowed you to emulate and imitate other people’s actions.” (Whatever you think of the plausibility of Rama’s idea, it’s worth a listen just to bask in the mellifluous rolling of his *rs*!)

In response to this embrace of mirror neurons as a factor accounting for important human behaviors, and their pathological dysfunction, cautionary voices have been raised. For these, the reader is directed to two opinion pieces, one entitled “A Mirror up to Nature” (Dinstein, Thomas, Behrmann, and Heeger, 2008), and a second entitled “Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans” (Hickok, 2009), both listed under *Further Reading*.

END-OF-CHAPTER QUESTIONS

1. What is the principle of efference copy? Name two major systems that work on this principle.
2. What does it mean to say that muscles behave like springs? How does the motor system take advantage of this property?
3. In what way(s) is the motor homunculus similar to the somatosensory homunculus? In what way(s) are they different?
4. Explain how population coding across hundreds of broadly tuned units can produce exquisitely precise movements.
5. What principle do you think better captures the function of the parietal cortex, processing *where* objects are located in the visual world, or computing *how* to act on the world based on visual input? Or, alternatively, can both be valid? Support your answer with empirical evidence.
6. How is a copy of the efferent motor signal processed in the cerebellum? What is the result?
7. What are two consequences for basal ganglia functioning of reducing levels of DA in the neostriatum, as occurs in Parkinson's disease?
8. What property(properties) distinguish(es) mirror neurons of premotor cortex from "conventional" neurons in this region. What functions might such neurons enable?

REFERENCES

- Albus, James S. 1971. "A Theory of Cerebellar Function." *Mathematical Biosciences* 10 (1–2): 25–61. doi: 10.1016/0025-5564(71)90051-4.
- Alexander, Garrett E., and Michael D. Crutcher. 1990. "Functional Architecture of Basal Ganglia Circuits: Neural Substrates of Parallel Processing." *Trends in Neuroscience* 13 (7): 266–271. doi: 10.1016/0166-2236(90)90107-L.
- Alexander, Garrett E., Mahlon R. DeLong, and Peter L. Strick. 1986. "Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex." *Annual Review of Neuroscience* 9: 357–381. doi: 10.1146/annurev.neuro.9.1.357.
- Andersen, Richard A., Sam Musallam, and Bijan Pesaran. 2004. "Selecting the Signals for a Brain–Machine Interface." *Current Opinion in Neurobiology* 14 (6): 720–726. doi: 10.1016/j.conb.2004.10.005.
- Brown, Peter, Antonio Oliviero, Paolo Mazzone, Angelo Insola, Pietro Tonali, and Vincenzo Di Lazzaro. 2001. "Dopamine Dependency of Oscillations between Subthalamic Nucleus and Pallidum in Parkinson's Disease." *Journal of Neuroscience* 21 (3): 1033–1038.
- Carmena, Jose M., Mikhail A. Lebedev, Roy E. Crist, Joseph E. O'Doherty, David M. Santucci, Dragan F. Dimitrov, ... Miguel A. L. Nicolelis. 2003. "Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates." *PLoS Biology* 1 (2): e42. doi: 10.1371/journal.pbio.0000042.
- Cohen-Séat, Gilbert, Henri Gastaut, J. Faure, and Georges Heuyer. 1954. "Études expérimentales de l'activité nerveuse pendant la projection cinématographique." *Revue internationale de filmologie* 5: 7–64.
- Dapretto, Mirella, Mari S. Davies, Jennifer H. Pfeifer, Ashley A. Scott, Marian Sigman, Susan Y. Bookheimer, and Marco Iacoboni. 2005. "Understanding Emotions in Others: Mirror Neuron Dysfunction in Children with Autism Spectrum Disorders." *Nature Neuroscience* 9 (1): 28–30. doi: 10.1038/nn1611.
- Eccles, John C., János Szentágothai, and Masao Itô. 1967. *The Cerebellum as a Neuronal Machine*. Berlin: Springer-Verlag.
- Gallese, Vittorio, Luciano Fadiga, Leonardo Fogassi, and Giacomo Rizzolatti. 1996. "Action Recognition in the Premotor Cortex." *Brain* 119 (2): 593–609. doi: 10.1093/brain/119.2.593.
- Gastaut, Henri J., and J. Bert. 1954. "EEG Changes During Cinematographic Presentation." *Electroencephalography and Clinical Neurophysiology* 6 (3): 433–444. doi: 10.1016/0013-4694(54)90058-9.
- Georgopoulos, Apostolos P., Roberto Caminiti, John F. Kalaska, and Joe T. Massey. 1983. "Spatial Coding of Movement: A Hypothesis Concerning the Coding of Movement Direction by Motor Cortical Populations." *Experimental Brain Research Supplementum* 7: 327–336. doi: 10.1007/978-3-642-68915-4_34.

- Georgopoulos, Apostolos P., John F. Kalaska, Roberto Caminiti, and Joe T. Massey. 1982. "On the Relations between the Direction of Two-Dimensional Arm Movements and Cell Discharge in Primate Motor Cortex." *Journal of Neuroscience* 2 (11): 1527–1537.
- Goodale, Melvyn A., A. David Milner, Lorna S. Jakobson, and David P. Carey. 1991. "A Neurological Dissociation between Perceiving Objects and Grasping Them." *Nature* 349 (6305): 154–156. doi: 10.1038/349154a0.
- Ito, Masao. 2008. "Control of Mental Activities by Internal Models in the Cerebellum." *Nature Reviews Neuroscience* 9 (4): 304–313. doi: 10.1038/nrn2332.
- Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell, eds. 2000. *Principles of Neural Science*. New York: McGraw-Hill.
- Kilavik, Björg Elisabeth, Joachim Confais, Adrián Ponce-Alvarez, Markus Diesmann, and Alexa Riehle. 2010. "Evoked Potentials in Motor Cortical Local Field Potentials Reflect Task Timing and Behavioral Performance." *Journal of Neurophysiology* 104 (5): 2338–2351. doi: 10.1152/jn.00250.2010.
- Penfield, Wilder, and Theodore Rasmussen. 1950. *The Cerebral Cortex of Man. A Clinical Study of Localization of Function*. New York: Macmillan.
- Pohl, Walter. 1973. "Dissociation of Spatial Discrimination Deficits Following Frontal and Parietal Lesions in Monkeys." *Journal of Comparative and Physiological Psychology* 82 (2): 227–239. doi: 10.1037/h0033922.
- Umiltà, Maria Alessandra. 2000. "Premotor Area F5 and Action Recognition." PhD diss., University of Parma.

OTHER SOURCES USED

- Albus, James S. 1989. "The Marr and Albus Theories of the Cerebellum: Two Early Models of Associative Memory." Paper presented at COMPCON Spring '89. Thirty-fourth IEEE Computer Society International Conference: Intellectual Leverage, Digest of Papers. San Francisco, California, February 27–March 1: 577–582. doi: 10.1109/COMPCON.1989.301996.
- Blakeslee, Sandra. 2008. "Monkey's Thoughts Propel Robot, a Step That May Help Humans." *The New York Times*, January 15. http://www.nytimes.com/2008/01/15/science/15robo.html?pagewanted=all&_r=0.
- Dinstein, Ilan, Cibu Thomas, Marlene Behrmann, and David J. Heeger. 2008. "A Mirror up to Nature." *Current Biology* 18 (1): R13–R18. doi: 10.1016/j.cub.2007.11.004.
- Georgopoulos, Apostolos P., Andrew B. Schwartz, and Ronald E. Kettner. 1986. "Neuronal Population coding of Movement Direction." *Science* 233 (4771): 1416–1419.
- Goodale, Melvyn A., and A. David Milner. 1992. "Separate Visual Pathways for Perception and Action." *Trends in Neurosciences* 15 (1): 20–25. doi: 10.1016/0166-2236(92)90344-8.
- Jakobson, Lorna S., Y. M. Archibald, David P. Carey, and Melvyn A. Goodale. 1991. "A Kinematic Analysis of Reaching and Grasping Movements in a Patient Recovering from Optic Ataxia." *Neuropsychologia* 29 (8): 803–805. doi: 10.1016/0028-3932(91)90073-H.
- Leveratto, Jean-Marc. 2009. "La revue internationale de filmologie et la genèse de la sociologie du cinéma en France." *Cinemas: revue d'études cinématographiques* 19 (2–3): 183–215. doi: 10.7202/037553ar.
- Mink, J. W., and W. T. Thach. 1991. "Basal Ganglia Motor Control. I. Nonexclusive Relation of Pallidal Discharge to Five Movement Modes." *Journal of Neurophysiology* 65 (2): 273–300.
- Tononi, Giulio. 2008. "Consciousness as Integrated Information: A Provisional Manifesto." *Biological Bulletin* 215 (3): 216–242.
- Umiltà, Maria Alessandra, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, and G. Rizzolatti. 2001. "I Know What you Are Doing. A Neurophysiological Study." *Neuron* 31(1): 155–65.

FURTHER READING

- Dinstein, Ilan, Cibu Thomas, Marlene Behrmann, and David J. Heeger. 2008. "A Mirror up to Nature." *Current Biology* 18 (1): R13–R18. doi: 10.1016/j.cub.2007.11.004.
- A critique of the mirror neuron literature, from neurophysiological and functional perspectives, from two leading cognitive neuroscientists and their trainees.*

- Hickok, Gregory. 2009. "Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans." *Journal of Cognitive Neuroscience* 21 (7): 1229–1243. doi: 10.1162/jocn.2009.21189.
The title says it all.
- Hickok, Gregory. 2014. *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition*. New York: W. W. Norton.
An update and elaboration on the 2009 opinion piece that not only critiques the "action understanding" theory, but also reviews alternative proposals about mirror neuron contributions to motor control. Some of this book's content is also relevant to the work of Hickok, and others, that is detailed in Chapter 18: Language.
- Himmelbach, Marc, Rebecca Boehme, and Hans-Otto Karnath. 2012. "20 Years Later: A Second Look on D. F.'s Motor Behaviour." *Neuropsychologia* 560 (1): 139–144. doi: 10.1016/j.neuropsychologia.2011.11.011.
A comparison of more recently acquired control data with the original results from D. F. that concludes that "the widespread and popular presentation of strong dissociations between distinct visual systems seems to be exaggerated."
- Jenkinson, Ned, and Peter Brown. 2011. "New Insights into the Relationship between Dopamine, Beta Oscillations, and Motor Function." *Trends in Neurosciences* 34 (12): 611–618. doi: 10.1016/j.tins.2011.09.003.
Review that nicely demonstrates how clinical research (in this case, with Parkinson's disease) can inform basic science-level knowledge, and vice versa.
- Kaas, Jon H. 2012. "Evolution of Columns, Modules, and Domains in the Neocortex of Primates." *Proceedings of the National Academy of Sciences of the United States of America* 109 (Supplement 1): 10655–10660. doi: 10.1073/pnas.1201892109.
Although some of this paper is focused on fine-grained anatomical details about cortical columns, it also advances the idea that classical somatotopy in primary motor cortex is "fractured," and that the fundamental organizing principle may be in terms of "complex movement sequences" that can be categorized as "reaching," "grasping," and "defensive" actions.
- Kravitz, Dwight J., Kadharbatcha S. Saleem, Chris I. Baker, and Mortimer Mishkin. 2011. "A New Neural Framework for Visuospatial Processing." *Nature Reviews Neuroscience* 12 (4): 217–230. doi: 10.1038/nrn3008.
A proposal that discrete anatomical pathways support the processing of spatial information in the service of where-based behavior (PPC-EEF and PFC), how-based behavior (PPC-premotor cortex), and navigation behavior (PPC-medial PPC-MTL).
- Rizzolatti, G., and L. Craighero. 2004. "The Mirror-Neuron System." *Annual Review of Neuroscience* 27: 169–192. doi: 10.1146/annurev.neuro.27.070203.144230.
This review has, at the time of this writing, accrued an astounding 3588 citations in the peer-reviewed literature.
- Schenk, Thomas, and Robert D. McIntosh. 2010. "Do We Have Independent Visual Streams for Perception and Action?" *Cognitive Neuroscience* 1(1): 52–62. doi: 10.1080/17588920903388950.
A "20-years later" critical review of the what vs. how model, followed by commentary/rebuttals from many leading researchers, including Goodale and Milner.

CHAPTER 8

OCULOMOTOR CONTROL AND THE CONTROL OF ATTENTION

KEY THEMES

- Four regions that contribute importantly to oculomotor control are the superior colliculus, regions of occipital and parietal cortex (particularly V1 and LIP), regions of frontal cortex (the frontal eye field (FEF) and the supplementary eye field (SEF)), and the brainstem oculomotor nuclei that send commands to the extraocular muscles.
- Although superficial layers of the superior colliculus receive direct projections from the retina, the deep layers contain multimodal sensory neurons and movement neurons to which posterior cortical neurons project, and which, in turn, project to the brainstem oculomotor nuclei.
- FEF neurons, which project directly to the brainstem oculomotor nuclei, encode vector saccades, whereas SEF neurons encode goal-oriented saccades.
- Saccade planning and the endogenous control of attention both engage a dorsal frontoparietal network that includes FEF and IPS/LIP.
- The reflexive, exogenous capture of attention is associated with a right hemisphere-biased ventral cortical system, including the temporal-parietal junction and the inferior PFC, regions also associated with hemispatial neglect.
- Microstimulation of monkey FEF produces retinotopically specific behavior that is equivalent to the effects of cuing the endogenous control of attention, and also produces attention-like effects in V4 neurons whose receptive fields overlap with the movement field of the stimulation site.

- Multivariate pattern analysis (MVPA) can dramatically increase the sensitivity of neuroimaging data analyses, and provide novel insights about information processing in the brain.
- MVPA has been used to demonstrate that the same systems that control eye movements are also used to control attention.

CONTENTS

KEY THEMES

ATTENTION AND ACTION

WHYS AND HOWS OF EYE MOVEMENTS

Three categories of eye movements

THE ORGANIZATION OF THE OCULOMOTOR SYSTEM

An overview of the circuitry

The superior colliculus

The posterior system

The frontal eye field

The supplementary eye field

THE CONTROL OF EYE MOVEMENTS, AND OF ATTENTION, IN HUMANS

Human oculomotor control

Human attentional control

The endogenous control of attention

The exogenous control of attention

THE CONTROL OF ATTENTION VIA THE OCULOMOTOR SYSTEM

Covert attention

Covert attention as motor preparation?

Empirical evidence for covert attention as motor preparation

Where's the attentional controller?

The reentry hypothesis

ARE OCULOMOTOR CONTROL AND ATTENTIONAL CONTROL REALLY THE "SAME THING"?

The "method of visual inspection"

"Prioritized Maps of Space in Human Frontoparietal Cortex"

OF LABELS AND MECHANISMS END-OF-CHAPTER QUESTIONS REFERENCES OTHER SOURCES USED FURTHER READING

ATTENTION AND ACTION

We have reached the final chapter of *Section II*, themed *Sensation, Perception, Attention, and Action*. Whereas we have seen in previous chapters how strongly perception and attention are linked (e.g., the attentional amplification of one sensory input channel over others, the dependence on attention of experience-dependent plasticity, unilateral neglect), this chapter will emphasize even more strongly than did the previous two how closely intertwined are the constructs of attention and action. At another level, this is the last chapter for which it will be necessary to begin with an overview of the anatomy of the system of interest. We haven't, of course, covered "everything," but upon completion of this chapter, the reader will have had an introduction to all of the major systems and principles of functional neuroanatomy, and will have the foundation on which to build a more detailed understanding of the intricacies of any particular component. Similarly, if the author has done his job, we'll see in every chapter following this one – i.e., in every chapter detailing some aspect of "high-level cognition" – that the principles underlying memory storage and retrieval, cognitive control, decision making, language, emotion, and so on, can be understood as variations on the principles introduced in this section on the more "elemental" functions of the brain.

WHYS AND HOWS OF EYE MOVEMENTS

Let's take a moment to recall one of the reasons why the ability to move the eyes is so important: it enables the visual system to disproportionately devote processing resources and, therefore, acuity, to just a small area of the visual field. If you need detailed information about an object located 15° to the left of your current point of fixation, you can obtain it within ~ 300 ms by moving your center of gaze to this location. The trade-off for the "cost" of those 300 ms is not having to drag around a visual cortex that's the size of the room that you're currently sitting in, or, alternatively, not having to survive with a visual system with uniformly horribly low resolution.

Because there are some instances when a few tens of milliseconds can matter, however, we'll see that one feature of the oculomotor system is multiple pathways that control the eyes in different contexts.

Three categories of eye movements

Eye movements fall into three categories: smooth pursuit, vergence shifts, and saccades. We won't spend much time considering the first two, but will define them briefly here. If you hold out your thumb and fixate the nail, then slowly move your head from side to side, all the while maintaining fixation on your thumbnail, you are performing smooth-pursuit eye movements. (If your eyes hadn't been moving to compensate for the back-and-forth motion of your head, your thumbnail wouldn't have remained in your center of gaze.) The equivalent eye movements are generated if you now hold your head still and move your thumb back and forth. Thus, smooth-pursuit eye movements enable you to maintain fixation on an object while it, you, or both, are moving. During smooth pursuit, the two eyes move "together" (i.e., the movements of one eye are identical to those of the other). These are referred to as **conjugate eye movements**.

Vergence shifts occur when an object of interest is moving toward or away from you, and in order to fixate it with both eyes, the two must move in opposite directions: toward each other if the object is looming; away from each other if it is receding (*Tip 8.1*).

The third category, saccadic eye movements, is the focus of this chapter. The word *saccade* is French for a jerky movement, and was first associated with eye

Tip 8.1

Signals from stretch receptors on the extraocular muscles convey information about the positions of the eyes relative to each other (i.e., how parallel vs. converging). This information contributes to the perception of depth (i.e., seeing in 3-D), and of motion in depth (i.e., motion toward and away from you, as opposed to motion along a 2-D surface, such as when looking at a movie screen or a computer monitor).

movements by the French ophthalmologist Emile Javal (1839–1907), who held a mirror at an angle while reading and observed that his eyes made a series of brief fixations, punctuated by abrupt conjugate movements, rather than a smooth, continuous drift across the page as most (who had ever given it any thought) had previously assumed. Saccades are of primary interest here because of their central importance to many domains of cognition (reading being just one of them), and because they have been used extensively in neuroscience and psychology research as an indirect measure of cognitive operations. Because monkeys cannot be verbally instructed how to respond on experimental tasks, nor can they tell you what they are thinking, saccades are a way for the animal to indicate its choice in an experiment. (E.g., *Which is the most recent item that you saw? or Which of these stimuli is different from the others?*) And although it would be disingenuous to suggest that training a monkey to perform a cognitive task is ever “easy,” it is the case that saccadic eye movements are natural behaviors for monkeys (and humans), and their use as a dependent measure can at least give the researcher a “head start” in the training process. For humans, the saccade itself can be of fundamental interest, if it is a fundamental part of the behavior of interest (e.g., reading, or visual search). Additionally, scientists have devised clever ways to use eye movements to assay the dynamics of cognitive operations as they are being carried out, as opposed to, or in addition to, retrospective reports that come at the end of each trial. (E.g., whether the eyes start to drift toward word X or word Y, both displayed on a screen, can provide insight into the mental operations engaged while a person is listening to a semantically ambiguous sentence.)

THE ORGANIZATION OF THE OCULOMOTOR SYSTEM

An overview of the circuitry

There are many factors that can prompt an eye movement. I can decide, while taking a pause as I type this sentence, to look up at a painting hanging on the wall of my living room. My eyes can move reflexively toward the source of a sound, or to a touch (the cat

walks into the room, then jumps up onto the couch next to me). I can scan the shelves of a grocery store, looking for that particular brand of soy sauce. Much of the discussion here, however, will be framed in terms of visually guided saccades. That is, what are the steps that transpire between the detection on the retina of a visual stimulus, and the saccade that brings that stimulus into foveal vision. Along the way, we’ll note instances where the principles may be similar or different for these other factors that can also trigger saccades.

Although the principal pathway for visual processing in most mammalian species is the retino-geniculostriate pathway, it “wasn’t always that way.” In fish and amphibians, the primary site of visual processing is a midbrain structure known as the optic tectum (*Tip 8.2*). In the mammal, the homologue is the superior colliculus (“higher hill” in Latin; the inferior colliculus of the auditory system, *Figure 4.4*, is the “lower hill”), a structure that sits immediately caudal to the thalamus on the “roof” (“tectum”) of the brainstem. There is a direct, monosynaptic pathway from the retina to the superior colliculus, which will receive more consideration further along in this chapter. The superior colliculus projects to the **brainstem oculomotor nuclei**, which house the local circuitry

Tip 8.2

The implication here is that modern fish/amphibians and mammals share a common ancestor for which, as continues to be the case for fish/amphibians, the optic tectum was the sole brain structure that interceded between retina and brainstem oculomotor nuclei. This does not imply, however, that the optic tectum of modern fish and amphibians remains unchanged, “frozen in time,” from that of the common ancestor. Indeed, the optic tectum of many modern aquatic species is much more complicated than the mammalian superior colliculus – subsequent evolutionary changes in the visual systems of some species have been most pronounced in the optic tectum, and in those of others, in cerebral cortex.

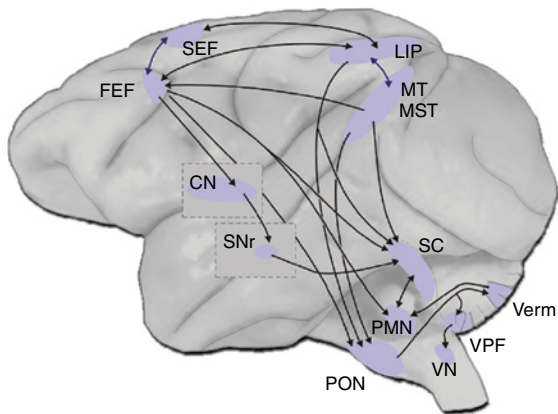


FIGURE 8.1 Important nodes and connections for oculomotor control. Note that among this diagram's simplifications are the absence of the V1 projection to SC, and the integration of LIP and FEF signals in the "oculomotor" cortico-basal ganglia-thalamic loop. Abbreviations: CN=caudate nucleus; FEF=frontal eye field; LIP=lateral intraparietal area; MT=middle temporal area; MST=medial superior temporal area; PMN=brain stem (pre)motor nuclei (PPRF, rMLF, cMRF); PON=precerebellar pontine nuclei; SC=superior colliculus (intermediate and deep layers); SEF=supplementary eye field; SNr=substantia nigra pars reticulata; Verm=oculomotor vermis (cerebellum, lobules VI and VII); VN=vestibular nuclei; VPF=ventral paraflocculus (cerebellum). Source: Krauzlis, 2005. Reproduced with permission of SAGE Publications.

and oculomotor neurons, analogous to the motor circuitry of the spinal cord (*Chapter 7*), that process descending motor commands and send axons that innervate the extraocular muscles. Within cortex, regions contributing to oculomotor control can be divided into posterior and anterior systems, the posterior including, most prominently, V1, LIP, and V4, and the anterior comprised of the frontal eye field (FEF) and the supplementary eye field (SEF, also referred to as the medial eye field (MEF) or dorsomedial frontal cortex (DMFC)). The primary distinction between the posterior and anterior systems is that the former regions project to the superior colliculus – superior colliculus lesions abolish the ability of electrical stimulation of posterior cortical areas to elicit saccades – whereas the anterior regions, although sending projections to the superior colliculus, also send projections directly to the brainstem oculomotor nuclei (*Figure 8.1*).

Tip 8.3

Keep in mind that most, if not all, of the cortical areas to be discussed here can't be considered "specialized" for oculomotor control; the label "oculomotor system" is simply an economical way to say "regions in the brain that can contribute to oculomotor control."

Let's now consider each of these major elements of the "oculomotor system" in turn (*Tip 8.3*).

The superior colliculus

In primates, the superior colliculi (one on each side of the midline) are each about the size of a half-globe of a pea, and roughly divided into three layers: superficial, intermediate, and deep. It is the deep layers that send efferents to the brainstem oculomotor nuclei. The superficial layers receive direct input from the retina, via a specialized class of retinal ganglion cells called *w cells*. The deep layers are of greater interest, because they receive the cortical efferents illustrated in *Figure 8.1*, as well as sensory information from the visual, auditory, and tactile modalities. There are two interleaved populations of cells: sensory neurons that have multimodal receptive fields and represent locations in retinotopic coordinates; and saccade neurons whose bursting activity brings the center of gaze into the receptive field of adjacent sensory neurons. Thus, the superior colliculus employs a vector code, in that activity in a certain area always produces a saccade of the same direction and magnitude (*Figure 8.2*; *Tip 8.4*).

At the rostral-most end of the superior colliculus, in the deep layers, are fixation neurons. In effect, in relation to the motor map illustrated in *Figure 8.2.B*, these command saccades of 0°. These neurons maintain a tonically elevated firing rate while one is fixating a stationary location. (Thus, at least from the reductionist perspective of activity in the superior colliculus, the term "passive fixation" that one frequently sees in the literature is a misnomer.)

At any moment in time, most of the motor units in the deep layers of the superior colliculus are under tonic inhibition from the substantia nigra pars reticulata, an arrangement analogous to the tonic inhibition of thalamus by the GPi that we detailed in *Chapter 7*.

FIGURE 8.2 The functional organization of the superior colliculus.

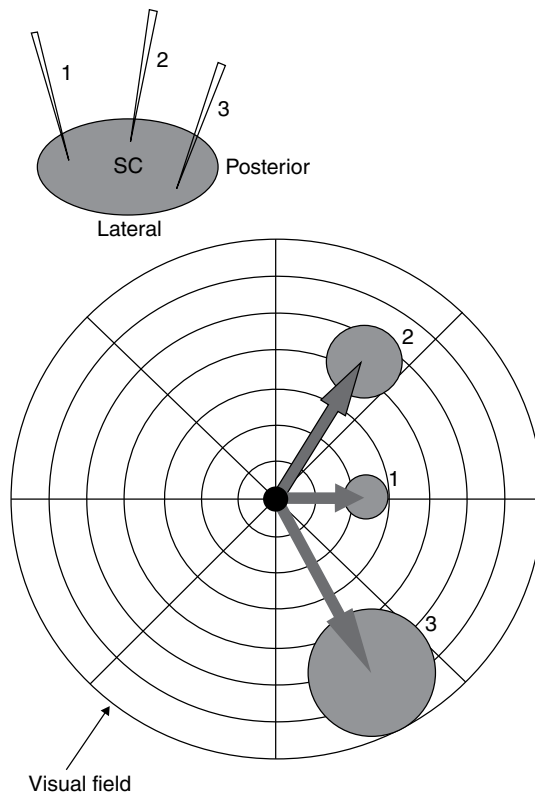


FIGURE 8.2.A Cartoon of the effects of recording from and stimulating at three electrode locations in the superior colliculus in the left hemisphere. The gray circles are (visual) sensory receptive fields, and the arrows indicate the magnitude and direction of the saccade generated when the neuron is stimulated. Source: Schiller, 1998. Reproduced with permission of Lawrence Erlbaum Associates.

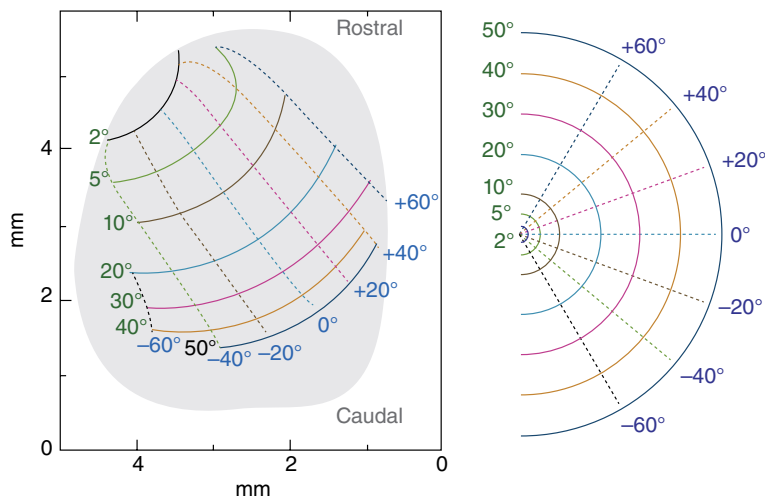


FIGURE 8.2.B Left panel: motor map in deep layers of the left superior colliculus; right panel, corresponding regions of the right visual field. Rostral areas represent small-amplitude saccades, caudal areas large amplitude saccades. Rostral and lateral to the 2° line (i.e., in the upper-left corner of the image) are fixation neurons. See Gandhi and Katnani, 2011. Source: Gandhi and Katnani, 2011. Reproduced with permission of Annual Reviews.

Tip 8.4

There is a potential for terminological confusion here, in that one might see an analogy to the cochlear nucleus, where “place code” refers to the fact that activity from one place along the cochlear nucleus always corresponds to a particular frequency; in the superior colliculus, activity at one place always produces saccades of the same metrics. In motor systems, however, the word “place” is used to refer to a location in the outside world. See *Figure 8.3*.

Tip 8.5

The property of “modulation” is traditionally ascribed to neurotransmitters ascending to cortex from subcortical nuclei (e.g., dopamine (see *Chapter 7*), norepinephrine, serotonin, and acetylcholine). However, Sherman and Guillery (2006) have proposed that many cortical glutamatergic neurons also play a modulatory, as opposed to a “driving” or an “information carrying,” role (more detail at *Web Link 8.1*).

The posterior system

Not too much ink needs to be spilt here, because the oculomotor functions of LIP were covered in some detail in *Chapter 6*. Although neurons from layer 5 of V1 send projections to the superior colliculus, their contribution seems to be modulatory, rather than driving. This assessment comes from the fact that, whereas microstimulation of V1 can produce saccades, the currents required to do so are markedly higher than those required in other cortical areas, such as the FEF, or in the superior colliculus itself. However, cooling V1, a method that reversibly inactivates a region of cortex, has the effect of abolishing visual sensory responses in the deep layers (*Tip 8.5*). Also consistent with a modulatory function of visual cortex on the superior colliculus is the association of oscillatory state in occipital cortex with the generation of **express saccades**. (For a more detailed treatment of the distinction between “driving” vs. modulatory signals, see *Web Link 8.1*.)

Express saccades have a latency that can be as much as one half that of standard saccade latency – in one recent study in humans, 90–135 ms after target onset for express saccades, 145–240 for standard saccades – and are often seen in response to the abrupt appearance of an unexpected stimulus. They are also readily produced experimentally with a “gap” paradigm, in which a gap of a few hundred ms is inserted between the offset of the fixation point that began the trial and the onset of the target to which the subject must make a saccade. Although it would be tempting to postulate that express saccades are generated by a

“short circuit” from retina to superior colliculus to brainstem oculomotor nuclei, the anatomy would seem to argue against this, because, as far as we know, retinocollicular projections terminate exclusively in superficial layers of the superior colliculus, and these do not project directly to the deep layers. An alternative account is that “advanced motor preparation of saccadic programs” (Paré and Munoz, 1996, p. 3666) enables the cortical system to generate saccades more quickly than is typical. Consistent with this idea is the fact that express saccades are more likely to be generated on trials with elevated oscillatory synchrony in the low alpha-band range, measured at occipital EEG electrodes.

Consistent with their role in processing visual motion, the contributions of areas MT and MST to oculomotor control are particularly important for guiding smooth pursuit movements.

The frontal eye field

As illustrated in *Figure 8.1*, the FEF in the monkey is located in the rostral bank of the arcuate sulcus, immediately anterior to premotor cortex. Among the first to study the FEF was David Ferrier, whose involvement in nineteenth-century research on the neural bases of vision made an appearance in *Chapter 1*. Ferrier (1875) found that electrical stimulation of this region produced “turning of the eyes and head to the opposite side” (p. 424), and that a unilateral lesion impaired the abilities of the animal to orient to stimuli presented contralateral to the lesion. Based on these and other observations, he stated most presciently in 1890 that “my hypothesis is that the

power of attention is intimately related to the volitional movements of the head and eyes” (Ferrier, 1890, p. 151). Consistent with this idea has been the subsequent description, in monkeys and humans, of unilateral neglect after damage to the FEF.

As indicated earlier, a distinguishing characteristic of the FEF is its low threshold for saccade generation with microstimulation. Although its organization isn’t strictly retinotopic, it’s generally the case that small, contralaterally directed saccades are produced by microstimulation in the ventrolateral limb of the anterior bank of the arcuate sulcus, larger amplitude saccades by microstimulation in the dorsomedial limb. This region encodes vector saccades, in that repeated stimulation of the same site produces a saccade of the same metrics, regardless of the starting point of the eye (*Figure 8.3*). A sustained pulse of stimulation produces a series of discrete saccades, all of the same metric. This is an interesting phenomenon that illustrates the recoding that happens in the brainstem oculomotor nuclei, because sustained microstimulation of a single location in one of these nuclei produces a very different effect – a saccade whose magnitude is linearly dependent on the duration of the stimulation

train. (The same is true, of course, for the extraocular muscles: they will only contract so long as ACh is being released into the neuromuscular junction.)

The supplementary eye field

Located dorsal and somewhat caudal to the FEF (*Figure 8.1*), the SEF lies just rostral to the SMA (*Chapter 7*). Its functions are associated with more “cognitive” aspects of oculomotor control. For example, when two targets are presented, and the monkey will receive a reward for saccading to either one (i.e., its choice), SEF, FEF, and LIP all show activity related to the impending choice, but that of the SEF starts earliest and is most robust. SEF activity is also prominent during the **antisaccade task**, when the rule is to make a saccade 180° opposite to the location of the target. The coding principle is also different between SEF and FEF, as is illustrated in *Figure 8.3*. Whereas microstimulation of the FEF neuron produces a saccade with roughly the same metrics regardless of the starting point of the eyes, microstimulation of a SEF neuron always produces a saccade that lands in the same “termination zone.” This is reminiscent of what

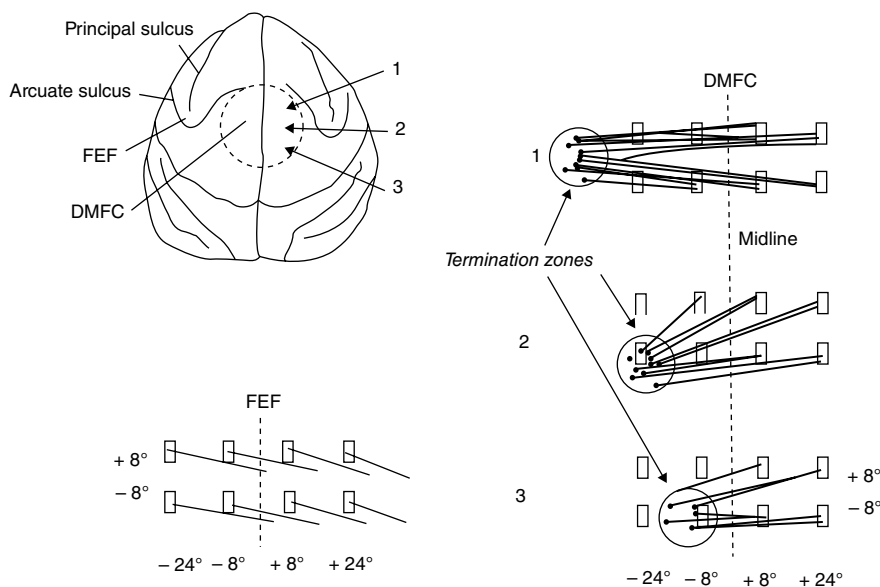


FIGURE 8.3 The effects of microstimulation of one neuron in FEF and of three neurons in SEF (referred to as DMFC by these authors), the location of each indicated with an arrow, from each of eight starting fixation points. Source: Schiller, 1998. Reproduced with permission of Lawrence Erlbaum Associates.

we saw in the spinal cord in *Chapter 7*, where stimulation of one location will always bring the forepaw to the same location in space, regardless of its starting location. This also highlights the terminological ambiguity referenced in *Tip 8.4*: a functional principle that has been ascribed to the SEF is a “place code,” in that activity in one set of neurons always brings the center of gaze to the same place. (Though, recall from *Chapter 7* that the similar phenomenon in rostral LIP was referred to as “goal-directed.”)

THE CONTROL OF EYE MOVEMENTS, AND OF ATTENTION, IN HUMANS

Human oculomotor control

A classic task for assessing oculomotor control is the oculomotor delayed-response task, in which the subject begins by fixating a central location, the to-be-acquired location is cued by a flashed stimulus, but the subject must withhold the saccade to that location until the offset of the fixation point. This temporal separation allows for the dissociation of “sensory” activity related to presentation of the cue from motor activity prompted by fixation offset. (The interpretation of “delay-period activity” sometimes observed between cue offset and fixation offset is less straightforward, as we shall see further along in this chapter, and again in *Chapter 13*.) *Figure 8.4* illustrates this task, together with examples of activity measured from the monkey and the human. Note that in the human, there is robust delay-period activity in the intraparietal sulcus (IPS; a superior portion of which is generally thought to be homologous to LIP in the monkey), as well as at two distinct frontal foci (*Figure 8.4.C*). The human FEF has classically been associated with the more dorsal of the two, just rostral and ventrolateral to the point of intersection of the precentral sulcus and the superior frontal sulcus. However, the more ventral focus of elevated activity in frontal cortex is also frequently observed. Note that although the frontal regions illustrated in *Figure 8.4.C* show elevated activity during each epoch of the trial (*cue, delay, response*), these regions

were identified statistically as having elevated activity during the delay period, regardless of whether or not they were differentially responsive during the *cue* and/or *response* epochs. Doing so requires an event-related procedure, some principles of which are described in *Methodology Box 8.1*, and in *Web Clip 3.2* and *Web Clip 8.1*.

The advent of event-related designs for fMRI seems to have spontaneously, independently occurred more-or-less at the same time in many laboratories in the late 1990s, including (from east to west) at the Functional Imaging Laboratory (FIL) of the University College London (UK), the Massachusetts General Hospital (Boston), the University of Pennsylvania (Philadelphia), the National Institutes of Health (Bethesda, MD), and Washington University in St. Louis (*Tip 8.7*). It is at the latter of these where we’ll next turn, to consider influential work on the control of attention.

Tip 8.6

The distinction between delayed response – in which the subject knows what the correct response is, but needs to withhold it until the “go” signal (e.g., the offset of the fixation point) – and delayed recognition – in which the response can’t be known until the presentation of the second stimulus – makes the former a test of motor preparation and the latter a test of short-term memory. (This distinction will be a focus of *Chapter 13*.)

Tip 8.7

The author remembers these as heady times for the cognitive neuroimaging community, finishing up his graduate studies, as he was, near one of these centers, and moving on to a postdoctoral fellowship at another. Buy him a drink sometime and he’ll tell you some stories.

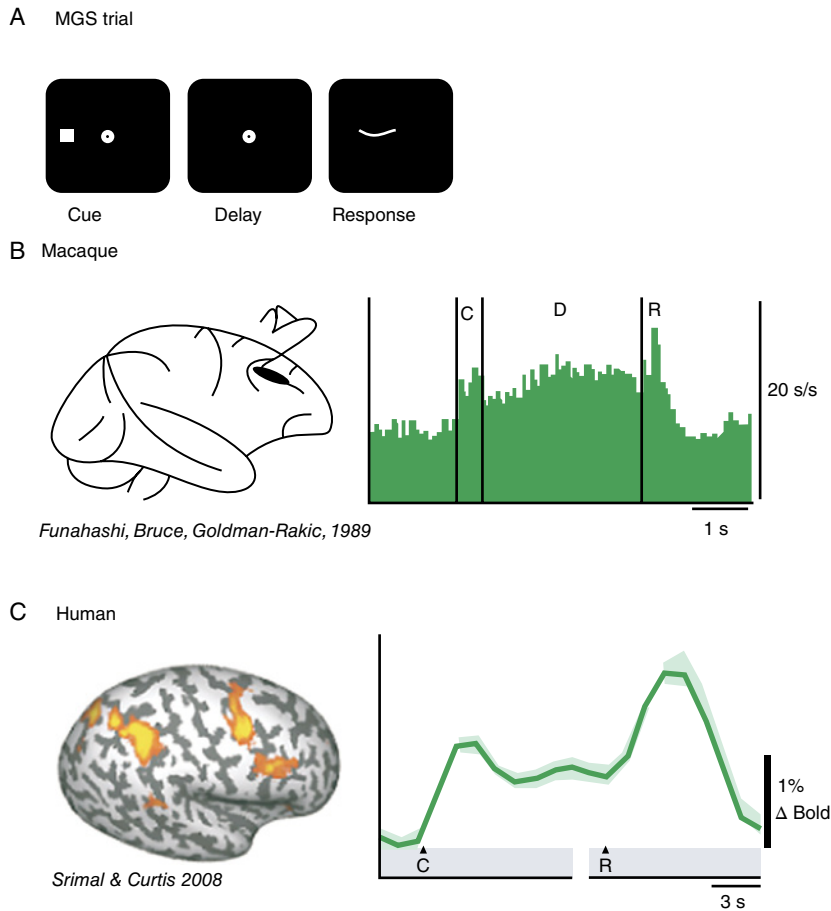


FIGURE 8.4 The delayed-saccade task (also referred to as “memory-guided saccade (MGS)” or “oculomotor delayed-response (ODR)”). Panel **A** illustrates a trial, as described in the text. Panel **B** illustrates the recording location in a monkey, in this instance at the caudal end of the principal sulcus, near the rostral boundary of the FEF, and activity recorded from a neuron at this location. Panel **C** illustrates regions with elevated delay-period activity in a human performing this task in an MRI scanner. The anatomical image has been “inflated,” so that sulci (darker gray) are visible. The fMRI time series data are interrupted between cue (C) and response (R) because the researchers used variable-length delay periods in this study. Source: Ikkai and Curtis, 2011. Reproduced with permission of Elsevier.

Human attentional control

Recall from *Chapter 6* that we restricted ourselves at that time to a detailed consideration of the *site* of the effects of attention, pushing off consideration of the *source* of attentional control to a later chapter. Well, *source*, your time has come.

From experimental psychology, we have long known that there are two general modes by which

attention can be directed: the exogenously triggered *capture* of attention vs. the endogenously controlled *allocation* of attention. An example of an exogenously triggered capture of attention would be that you are sitting quietly, reading this book, when suddenly someone loudly, unexpectedly, bursts into the room. Your reflexive reaction is to turn your head in the direction of the noise. An example of the endogenous control of attention is when you are sitting quietly, pretending to

METHODOLOGY BOX

8.1 Block designs vs. event-related designs in fMRI (Zarahn, Aguirre, and D’Esposito, 1997)

A challenge for both PET and fMRI studies is how to design tasks that allow for the isolation of a cognitive mechanism of interest. Recall from *Methodology Box 6.1* that the relatively low SNR of PET necessitates minutes-long periods of data acquisition. To isolate activity associated with speaking, therefore, the researchers subtracted images corresponding to blocks of time when subjects were viewing words from blocks from when they were reading the words aloud. Because the logic underlying this approach is that both blocks contain activity associated with the same cognitive operations (e.g., visually perceiving the letters, reading the word), except one (speaking the word aloud), it is called “cognitive subtraction.”

“Block designs,” in which several trials of the same type are blocked together, and the resultant neuro-imaging signal is averaged across the entire block, are also common with fMRI studies, as illustrated in *Figure 6.19*. One of the most compelling reasons for employing a block design is that it features higher SNR than event-related designs, by virtue of the inherent “sluggishness” of the hemodynamic response function (HRF; *Figure 6.8*). (The technical reasons for this are nicely explicated in *Further Reading* by Aguirre and D’Esposito (1999).) A downside of block designs, however, is that assumptions underlying the logic of cognitive subtraction can sometimes fail. *Figure 8.5*, for example, illustrates this with data from a *delayed-recognition* task (similar in some respects to the delayed-response task from *Figure 8.4*, but different in that subjects can’t know what the appropriate response is until the memory probe (labeled “discrimination/response”) is presented at the end of the trial (*Tip 8.6*)). *Figure 8.5.B* illustrates two hypothetical response patterns, one in which the brain region responds to stimulus presentation, but not during the delay period, the second in which the brain region responds during all three.

Figure 8.5.C illustrates the actual response of two adjacent regions of right PFC during this task, from Zarahn, Aguirre, and D’Esposito (1999). In these panels, we are privy to the moment-to-moment evolution of activity in these regions, because they were analyzed in an event-related manner. But let’s pretend for a moment that they were analyzed in a block design. Were there to have been, say, 10 consecutive D trials in a block, in both regions these would clearly have produced a significantly greater “area under the curve”

FIGURE 8.5 Event-related fMRI. Source: Zarahn, Aguirre, and D’Esposito, 1999. Reproduced with permission of Elsevier.

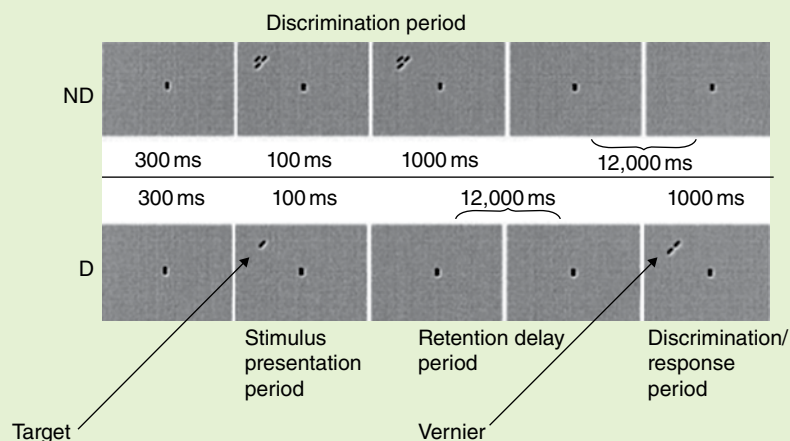


FIGURE 8.5.A The “no delay” (ND) and “delay” (D) variants of the task, in which the subject indicated whether the vernier array was further from, or closer to, fixation than the target bar.

Methodology Box 8.1 *Continued*

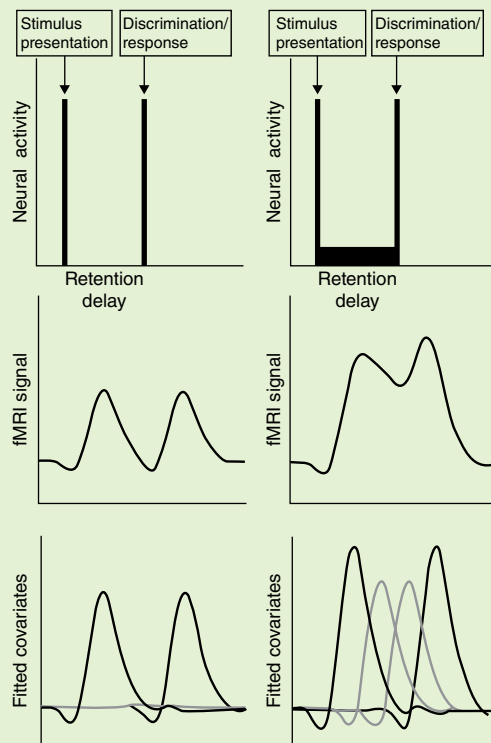


FIGURE 8.5.B A simulation of how two brain areas might respond during “D” trials on this task. The region in the left-hand column shows elevated activity during the “stimulus presentation” and “discrimination/response” epochs of the trial, but not during the delay period; the region in the right-hand column shows elevated activity across all three components of the trial. The middle row illustrates what the (noise-free) fMRI response would be in these two regions. The bottom row shows loadings on regressors (“fitted covariates”) in a regression model set up to analyze these event-related data.

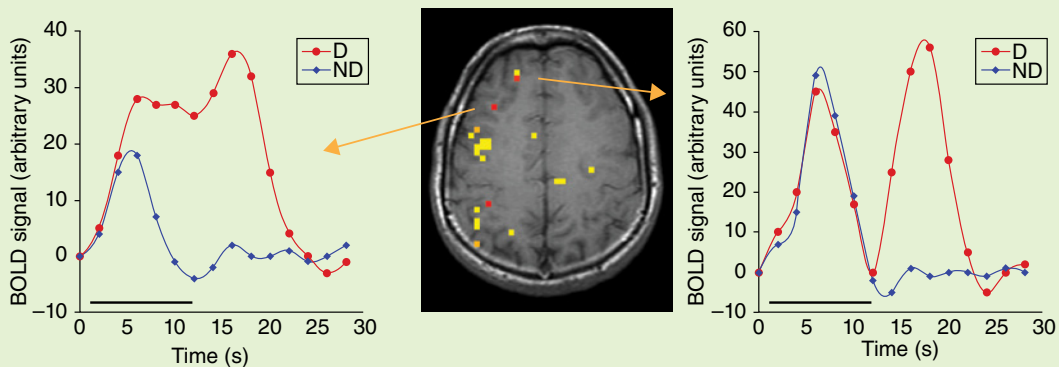


FIGURE 8.5.C The responses of two regions in the right PFC during ND and D trials. The black horizontal bar denotes duration of delay period. Note that, in the more anterior voxels, even though there is no elevated delay-period activity during D trials, discrimination/response activity is markedly higher in D trials than in ND trials. Thus, averaging across several trials of both types, as one would do in a block design, would yield a higher signal intensity value for these voxels for D trials than for ND trials, even though these voxels evince no elevated delay-period activity. Replotted by the author from Zarahn et al. (1999).

Methodology Box 8.1 *Continued*

value than a block of 10 consecutive ND trials. The interpretation, therefore, would be that both these regions exhibit significantly elevated delay-period activity. This is because the two trial types are matched with regard to the amount of time that stimuli are on the screen, the decision and response that are required, and so on. They only differ in that D trials include a delay period and ND trials do not. For the region plotted on the right, however, this interpretation would be incorrect. The activity of this region illustrates a failure of the assumption of “pure insertion,” in that insertion of a delay period did *not* leave the other epochs of the task unaffected. Specifically, “discrimination/response” activity increases when a delay is inserted. (Perhaps a memory-based decision is more difficult than a perceptually based one, leading to greater probe-related activity on D vs. ND trials?)

be reading this book, but are, in fact, waiting for the imminent arrival of your roommate/boyfriend/girlfriend/spouse/whomever, and so are attending to the door, waiting for it to open at any moment.

The endogenous control of attention

Researchers at Washington University in St. Louis, led by neurologist Maurizio Corbetta and experimental psychologist Gordon Shulman, have used event-related fMRI to separate, for example, transient responses associated with a cue instructing the subject where to attend (i.e., to a location in the periphery) – or what to attend to (e.g., motion in a particular direction) – from the sustained activity associated with the top-down maintenance of attention at the cued location – or to the cued direction of motion. Importantly, on perceptually identical “passive viewing” trials, when subjects were told that there was no task to perform, and that they merely had to watch the screen, the transient responses in posterior occipitotemporal regions were only slightly attenuated relative to what they had been in attention-demanding trials (*Tip 8.8*). This was consistent with what we have already seen at the site of attention. Activity in rostral IPS and caudal superior frontal cortex, including the FEF, however, showed a very different pattern.

Tip 8.8

In the human, MT is located on the lateral surface of the occipital lobe, at the dorsoventral level of the inferior temporal gyrus.

It was markedly attenuated, almost to baseline levels, on passive trials. Also noteworthy was the fact that virtually identical patches of cortex in superior parietal and frontal regions showed elevated activity regardless of whether the sustained attention required by a task was location-based (e.g., *Detect the onset of a target at this location*) or feature-based (e.g., *Indicate when you see motion in this direction*). This led to the conclusion, consistent with results from many, many other studies performed in many other laboratories, that coordinated activity in this dorsal frontoparietal network acted as a *source* of top-down attentional control (*Figure 8.6*).

How might the dorsal frontoparietal system work? At the network level, projections from deep layers of cortex in these regions likely deliver the signals that manifest in sensory processing areas as baseline shifts (in the case of location-based attention) and/or as the biasing signal that influences the competition for representation of a behaviorally relevant stimulus vs. irrelevant items. (Revisit *Chapter 6* if the phenomena of the baseline shift and of biased competition aren't fluently springing to mind.) And at a functional level, many researchers have proposed that elevated activity in dorsal frontoparietal cortex (*Figure 8.6*) may reflect the operation of a **saliency map**. The idea is that the brain maintains a representation of what's currently the most relevant for behavior, whether that be a representation of the location of potentially important stimuli in the environment, a representation of current behavioral priorities, or both.

Let's say you're new to Madison, WI, and are driving in a rental car from the airport, trying to navigate

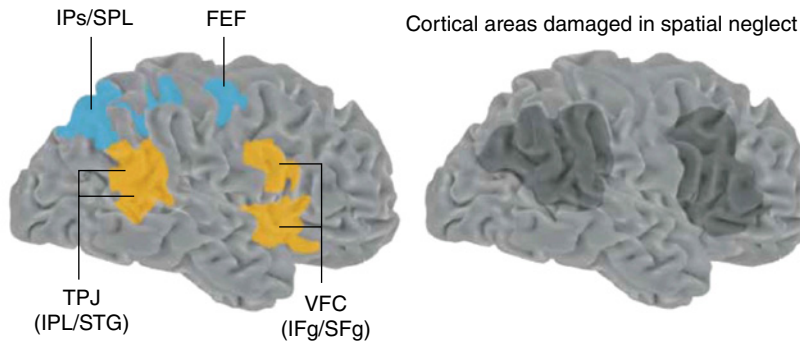


FIGURE 8.6 Two networks for the control of attention. At left, the networks associated with the endogenous (blue) vs. the exogenous (yellow) control of attention. At right, regions whose damage is often associated with hemispatial neglect. Source: Corbetta and Shulman, 2002. Reproduced with permission of Nature Publishing Group.

to my house. As you drive along Old University Avenue your eyes are, of course, on the road ahead of you. In your “mind” however, are additional representations: (1) you are searching for street signs at each intersection, so as to identify “Forest St.” (which should be coming up on the left); (2) you are aware of a group of people standing near a crosswalk in front of you (*What’s the law here? Do I have to stop and let them cross in front of me, or will they wait for a break in the traffic?*); and, (3) what’s up with the car behind you that is driving right up on your @&&? Thus, although your eyes are, indeed, on the road, your salience map will prioritize (1) the detection of green street signs; (2) movement into the crosswalk by a pedestrian; (3) evidence of trouble in the rearview mirror.

★★SUDDENLY, A BALL ROLLS OUT ONTO THE STREET HALFWAY DOWN THE BLOCK!★★

This unexpected event triggers what Corbetta and Shulman (2002) call the “circuit breaker,” which overrides the influence of your current salience map on your behavior.

The exogenous control of attention

Exogenous control of attention occurs when something salient and unexpected occurs in the world around you, and the source is, therefore, *outside* of your brain. (The endogenous control of attention, in contrast, is considered more “cognitive,” in the sense that your thoughts/plans/priorities/motivations are

Tip 8.9

Perhaps some of you, like me, now find it easier to reconcile *Chapter 6*’s account of unilateral neglect, and particularly Figure 6.4, with the rest of that chapter’s decided emphasis of more dorsal regions of PPC.

controlling your behavior. They are endogenous in the sense that they are generated within and maintained *inside* you, in your brain.) Corbetta and Shulman have documented that exogenous shifts of attention are associated with a right hemisphere–dominant ventral network that includes the temporoparietal junction (TPJ) and the inferior frontal gyrus of the ventral prefrontal cortex (which they abbreviate as VFC). The ventral network is illustrated in yellow in the brain image on the left of *Figure 8.6*. Intriguingly, Corbetta and Shulman (2002) have noted that “the anatomy of neglect ... matches the ventral TPJ–VFC system” (p. 212). (For one independent dataset generally consistent with this view (and generated several years later), revisit *Figure 6.4*; *Tip 8.9*.)

Next, we turn our attention to something that many readers will have already noticed, which is that the dorsal regions highlighted in panel C of *Figure 8.4*, which illustrates activity associated with a delayed-saccade task, show remarkable overlap with the dorsal frontoparietal system illustrated in *Figure 8.6*.

THE CONTROL OF ATTENTION VIA THE OCULOMOTOR SYSTEM

Covert attention

There are many times in our lives when we want to attend to something, or someone, without directly looking at it (or him, or her). In high school math class there was that cute classmate on whom you had a crush, but at whom you didn't want to stare (Never happened to me, but I've heard stories ...). At the dinner table, one wants to convey to one's kids that one trusts them, and so doesn't want to appear to be closely monitoring their every move, yet one also wants to closely monitor their every move (Again, so I'm told ...). These are examples of covert attention. (Overt attention is when you look directly at the object of your attention.) Sustained covert attention may be related to another attentional phenomenon, which is that there is considerable data from experimental psychology, and from neurophysiology, that the focus of attention jumps to the target of the next saccade tens of milliseconds prior to the initiation of the saccade itself. Therefore, might it be the case that the sustained allocation of covert attention to a location/object amounts to a presaccadic shift of attention, but one that does not culminate in the saccade itself?

Covert attention as motor preparation?

This last bit of speculation is essentially what is posited by the **premotor theory of attention**, articulated in the 1980s by Giacomo Rizzolatti (of mirror neuron renown). It argues, in summary, that attention may fundamentally be “nothing more” than motor planning [my quotes]. Here's how Rizzolatti and Craighero (2010) have summarized it:

Classically, spatial attention was thought of as a dedicated supramodal control mechanism, anatomically distinct from the circuits underlying sensorimotor processing ... [However] there is no need to postulate two control mechanisms, one for action and one for attention. According to [the premotor theory,] spatial attention does not result from a dedicated control mechanism, but derives from a weaker activation of the same frontal-parietal circuits that, in other conditions, determine motor behavior toward specific spatial locations.

Empirical evidence for covert attention as motor preparation

The first of a series of studies by Tirin Moore and colleagues that have put this theory to the test is described in *Research Spotlight 8.1*. Learning the details and implications of this experiment may bring one, in the humble estimation of this author, the closest that one could hope to get, in a textbook such as this, to Nabokov's “aesthetic bliss” (*Tip 8.10*).

A second study from Moore offers an even more direct demonstration of the premotor theory, and is even more technically breathtaking than what's described in *Research Spotlight 8.1*. Additionally, it shows that this scheme can generalize to object-based attention, such as investigated by Chelazzi, Duncan, Miller, and Desimone (1998) and detailed in *Chapter 6*. In it, Moore and graduate student Kathleen Armstrong (2003) matched up FEF neurons and V4 neurons whose motor fields and receptive fields, respectively, overlapped (*Tip 8.11*). In this way, they were able to stimulate at a putative source of attentional control while simultaneously recording from a site. *Figure 8.8* illustrates that the effect of FEF stimulation was to boost the V4 visual response in a manner similar to the effects of selective attention that we considered in *Chapter 6*. Importantly, in this study *the monkey was not performing a task*, other than maintaining central fixation. Thus, as with Moore and Fallah (2001), it is a remarkable demonstration of the ability to control the focus of attention via a decidedly exogenous intervention in the activity of a neural source of the endogenous control of attention (*Tip 8.12*)!

Where's the attentional controller?

Remember in *Chapter 7*, when a skeptic chimed up about “some little homuncular mathematician perched up on the precentral gyrus ... calculating vector sums prior to each movement”? At this point s/he might chime in again, asking sarcastically if we've placed the spotlight operator (from *Chapter 6*'s “Day at the Circus”) in the FEF. Stilted rhetorical devices aside, this is actually a legitimate concern, and one about which the cognitive neuroscientist and, particularly, the theoretician, needs always to be mindful. *Where does this control come from? Does assigning it to the dorsal frontoparietal system make it any less mystical, any more material?*

RESEARCH SPOTLIGHT

8.1 Controlling the spotlight of attention with microstimulation of the FEF (Moore and Fallah, 2001)

In previous chapters we have emphasized the value of being able to causally influence the system under study in systematic ways. This is illustrated very elegantly in the study highlighted here, which tests a hypothesized mechanism of attentional control. The first step of the experiment is to find a neuron in FEF whose microstimulation produces a repeatable saccade to the same location (e.g., *Figure 8.3*). This is referred to as the cell's motor field (by analogy to the sensory receptive field; panel A in *Figure 8.7*). Next, the layout of the covert attention task – on which the monkey has already been trained – is arranged so that the target stimulus is located in that neuron's motor field (panel B in *Figure 8.7*). In the task, the monkey is to depress

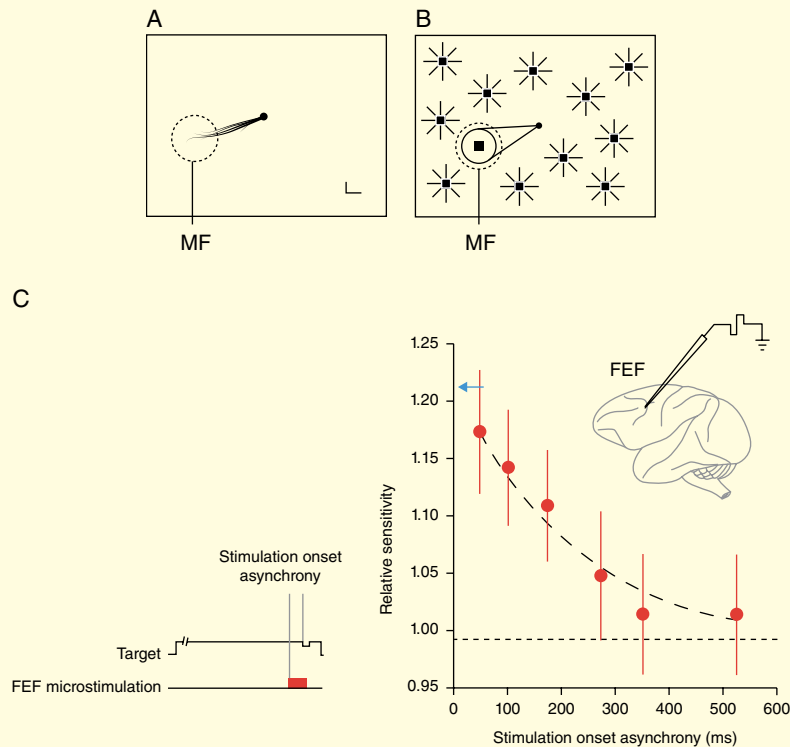


FIGURE 8.7 Electrically controlling the source of endogenous attentional control. **A.** Microstimulation of a location in the FEF of a monkey produces reliable vector saccades into a motor field (MF). **B.** The array of the visual search task is customized for that neuron, such that the target location falls within this MF. **C.** Microstimulation-related improvement in performance varied monotonically with the stimulation onset asynchrony (see text). Source: Awh, Armstrong, and Moore, 2006. Reproduced with permission of Nature Publishing Group.

Research Spotlight 8.1 *Continued*

a lever to start the trial, and maintain central fixation while attending to the target. When the target dims, ever so slightly, the monkey is to release the lever in order to receive its reward. The dimming of the target is made difficult to detect by making the luminance change very small and very brief, by making the timing of the event unpredictable, and by including catch trials in which the target never dims. Finally, as if that wasn't already enough, most trials also feature the concurrent presentation of flashing distractors throughout the trial. (Distractors flashed on and off, one at a time, in the locations illustrated with "sunbursts" in panel B in *Figure 8.7*.)

Now for the really cool part. Recall that the target is positioned within the motor field of the FEF neuron. This means that, if the investigator wanted to, he could deliver a 100 ms train of $\sim 50 \mu\text{A}$ of current and drive the eyes to the target location. Instead, they did what Tirin Moore sometimes characterizes as "just tickling" the neuron, by stimulating at half the amplitude that had been determined to produce a saccade 50% of the time. That is, the stimulation was subthreshold, in that it never produced an eye movement.

But would it – drumroll please – produce a covert shift of attention?

To assess this, the investigators varied the latency between when microstimulation was applied and when the target event occurred, a parameter called the "stimulation onset asynchrony." The findings suggest that "tickling" an FEF neuron does, indeed, produce a covert shift of attention, and that it does so in a timing-dependent manner: the closer in time the microstimulation was to the target event, the stronger were its attention-like effects. The effects, plotted in panel C in *Figure 8.7*, express the effect of microstimulation on the animal's "relative sensitivity," which is a measure of how much smaller the magnitude of the dimming needed to be made for the monkey to perform at the same level as it did on no-stimulation trials. Thus, in effect, what the plot shows is that delivering microstimulation with a stimulation onset asynchrony of 50 ms improved the animal's performance by roughly 17.5%. For comparison, the blue arrow shows that turning off the distracting stimuli improved the monkeys' performance by $\sim 21\%$.

In effect, instead of overtly cuing the animal about an impending target event, a la Posner (1980), Moore and Fallah "reached into its brain" and directly shifted the animal's attention!

Tip 8.10

For Nabokov, the best fiction afforded "a sense of being somehow, somewhere, connected with other states of being where art (curiosity, tenderness, kindness, ecstasy) is the norm" (Nabokov, 1955, p. 286). For me, the best science does this, too.

Have we accomplished anything more than assigning a more specific address to the ghost in the machine? The answer is that, in fact, in the domain of attentional control, and in many others, cognitive neuroscience has made important strides toward developing very explicit accounts of how control emerges as a consequence of

Tip 8.11

Tirin Moore had been a doctoral student of Charles Gross, of hand- and face-cell fame (*Chapter 5*), left to do a postdoc, then returned to Gross's lab (at Princeton University) for three additional years as a research scientist. Both Mazyar Fallah and Kathleen Armstrong were graduate students of Gross, collaborating with Moore, when this research was performed.

network interactions. In these models, there is no one place from which control signals are generated *de novo*, no figurative "corner office" from which a neural executive spontaneously decides to issue commands.

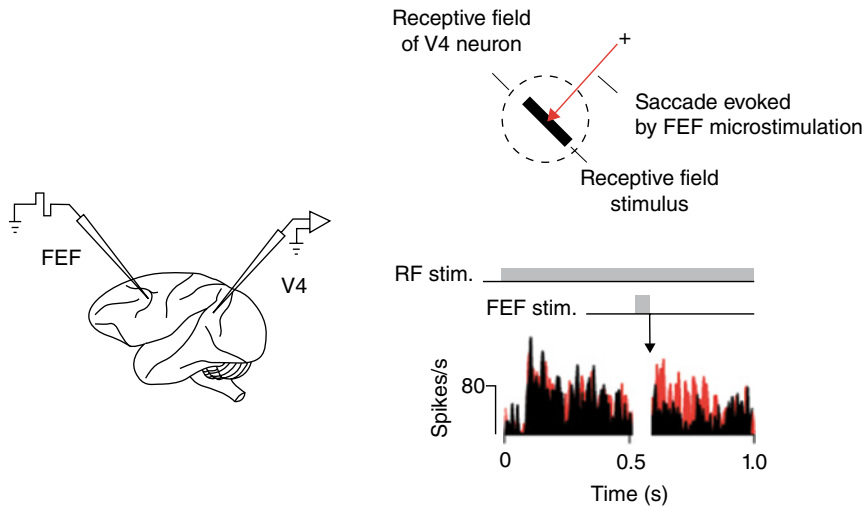


FIGURE 8.8 Generating the cause and measuring the effect of attentional control. Simultaneous stimulation of an FEF neuron and recording from a V4 neuron with overlapping motor field and receptive field, respectively. Effect of visual stimulus (oriented bar) on activity in the V4 neuron is illustrated on trials without FEF stimulation (black histogram) and with FEF stimulation (red histogram). Source: Awh, Armstrong, and Moore, 2006. Reproduced with permission of Nature Publishing Group.

Tip 8.12

If what you attend to determines what you think about, then what Moore and colleagues are doing may be, in effect, controlling what their monkeys are thinking about!

Such models are often most convincingly achieved via computational simulations. Although such simulations invariably include some idiosyncratic choices made by the modeler that may not account for every detail that we (think we) know about how the brain works, their value is in demonstrating that it is possible, in principle, to understand, to describe, and to measure every step, every connection, every causal factor in a system capable of producing complex behavior. For the question of the source of attentional control, we can consider a computational model by the German theoretician Fred Hamker.

The reentry hypothesis

Research Spotlight 8.2 details this computer network model, the goal of which was to simulate the findings of Chelazzi et al. (1998) that provide an important

empirical foundation for the biased-competition model of visual attention (*Chapter 6*). In particular, this model provides an explicit account for how the FEF (and IT) act as sources of the top-down attentional control that provide the biasing signal. Additionally, by specifying and explicitly modeling the activity of different classes of neurons in the FEF, it provides a plausible scenario for how a saccade plan might be prevented from actualization in a saccade, and, thus, could also serve as the basis for sustained covert attention. (Hint, it would involve the fixation neurons.)

Another clue to the question of *Who controls the controller?* comes from the research presented in *Web Clip 8.2*, which details the influence of the modulatory neurotransmitter dopamine on the attentional-control functions of the FEF.

ARE OCULOMOTOR CONTROL AND ATTENTIONAL CONTROL REALLY THE “SAME THING”?

So, we’ve seen evidence that FEF can act as a source of attentional control when it’s electrically stimulated, and we’ve considered a model describing how it

RESEARCH SPOTLIGHT

8.2 A computational account of the control of visual attention (Hamker, 2005)

Figure 8.9 illustrates the architecture of the model (panel B), and the anatomical connections on which it is based (panel A). One of the features of this model, which we haven't considered up to this point, is the heterogeneity of neuronal classes within FEF. Not unlike the deep layers of the superior colliculus, within FEF one can find visually responsive neurons, fixation neurons, visuomotor neurons, and movement neurons. Three of these are explicitly represented in the model (panel B in Figure 8.9), as are their temporal dynamics:

Visuomovement cells in deep layers are active from stimulus onset until saccade execution. Typically their initial response does not distinguish between distractor or target, but the activity decays when a distractor is in the receptive field. Movement cells are active prior to saccades and do not show any response to stimulus onset. Fixation cells decrease their activity before a planned eye movement. (Hamker, 2005, pp. 433–434)

Now, how does this network implement attentional control? We can watch it happen, in Figure 8.10, on a (simulated) millisecond-by-millisecond basis:

0 Not shown is the beginning of the trial, when initial presentation of the target object establishes the search template for that trial, which is retained in “working memory” in recurrent connections between PF and IT (between “Working memory” and “Target” populations) until the search array is presented. (For a cartoon of the task, see Figure 6.20.)

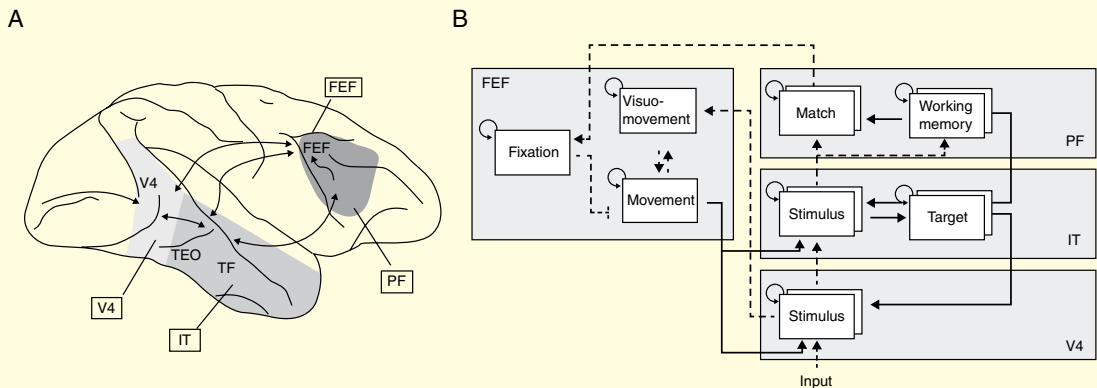


FIGURE 8.9 Illustrations from Hamker's simulation of the source of top-down control. Panel **A**, the known neural connections included in the model. Panel **B**, a schematic of the implementation of the model. Bottom-up connections are indicated in dotted arrows, feedback (“reentrant”) connections in solid arrows. Each white box represents a population of neurons, “stacked” boxes indicating populations processing two stimulus features (color and form) in parallel. The “turning back on itself” arrow at the upper-left corner of each of the boxes indicates that individual units within a population exert lateral inhibition on each other. Source: Hamker, 2005. Reproduced with permission of Oxford University Press.

Research Spotlight 8.2 *Continued*

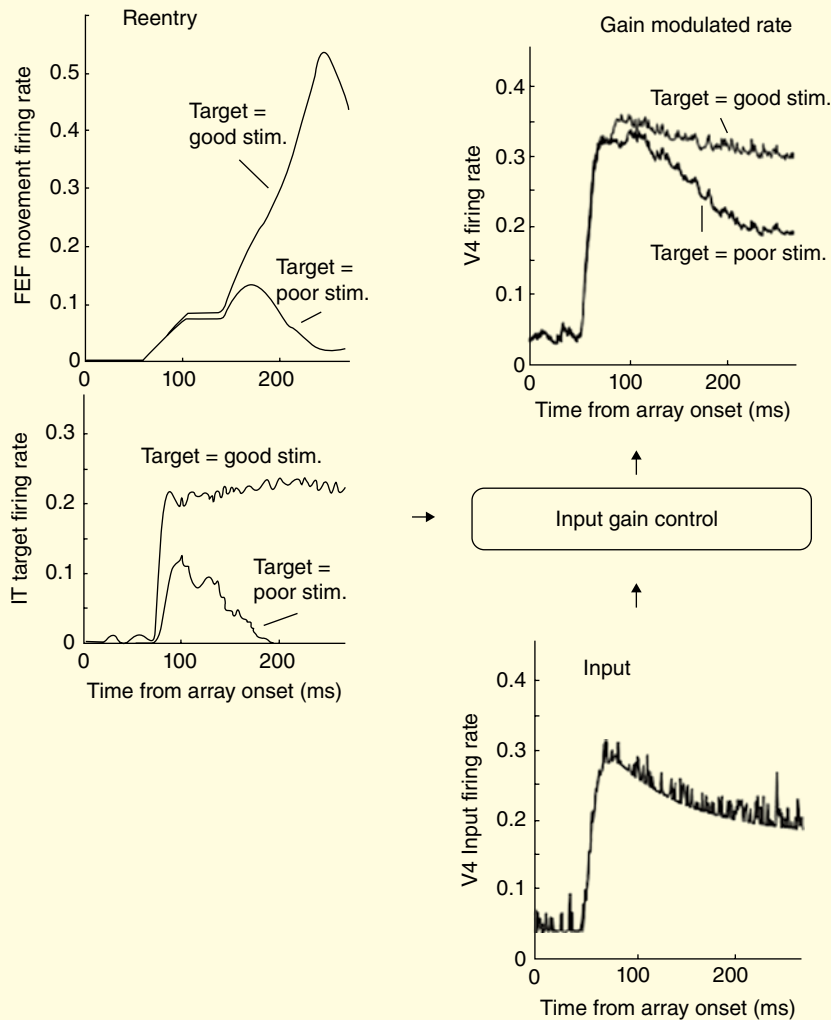


FIGURE 8.10 Multiple windows on the temporal evolution of activity occurring simultaneously at multiple stages of the simulated oculomotor system from *Figure 8.9*, for a V4 receptive field/FEF movement field processing information about a stimulus that is the V4 neuron's "good" stimulus (definitions in *Figure 6.20* and associated text). The "Input" box shows the stimulus-driven input to a V4 neuron from visual cortex. The "Input gain control" implements the **convolution** of this Input signal with the "Reentry" signals from FEF and IT (i.e., it's the operation in the model that computes the effects of the reentrant signals on the Input) and the "Gain modulated rate" shows the resultant activity in V4 (*Tip 8.13*). Source: Hamker, 2005. Reproduced with permission of Oxford University Press.

Research Spotlight 8.2 *Continued*

- 1 When one stimulus of the search array appears within the simulated receptive field, increased input elevates activity within V4 neurons tuned for particular stimulus features, and this feeds forward to FEF, producing the initial ramp-up in the *FEF movement firing rate* that begins ~ 75 ms into the trial.
- 2 Meanwhile, neurons in IT have knowledge (from working memory) about the target, and so their early activity to the stimulus reflects whether it is the target (*Target = good stim.*) or not.
- 3 [First wave of *reentry*:] Feedback from IT to V4 produces the initial little bump in *Target = good stim.* activity in the “Gain modulated rate” of V4 that begins prior to 100 ms.
- 4 Although the receptive field of any one V4 neuron is relatively large, the aggregate result of feedforward activity from a large population of them is that visuomovement neurons in FEF that encode the region of the visual field containing the target will receive the strongest feedforward drive (a phenomenon reminiscent of population coding in M1). The visuomovement neurons in FEF receiving the strongest feedforward drive will, in turn, activate the FEF movement neurons whose movement field overlaps the location of the target. The accumulating effect of the small differential feedforward drive from V4 begins to emerge in FEF activity at ~ 130 ms.
- 5 [Second wave of *reentry*:] Increasing activity in FEF movement neurons is experienced by V4 neurons as top-down feedback, and those representing features of the target, which already have a slightly elevated level of activity, benefit from this top-down boost to further increase their activity and, in turn, apply lateral inhibition to V4 neurons representing features of the nontarget stimulus.
- 6 These processes continue, in parallel, until a threshold is reached and a saccade to the target triggered.

(*Tip 8.14.*)

Tip 8.13

The Hamker (2005) paper presents several side-by-side comparisons of model outputs of simulated “Gain Modulated Rate,” from various conditions, with the comparable empirical data from Chelazzi et al. (1998; such as those illustrated in *Figure 6.21*).

might come to serve this function. But is there evidence that this *is* how attention is controlled in the human? That is, can the endogenous control of attention truly be reduced to “nothing more” than covert oculomotor control? Although this is a difficult question to answer unequivocally, we’ll conclude this chapter considering one way of addressing it.

Tip 8.14

Note that this model only takes into account firing rates, and how an increase (or decrease) in an area affects the region(s) that it projects to. There is, however, a growing body of electrophysiological evidence that dynamic changes in oscillatory phase synchrony between regions also plays an important role. *Web Clip 2.3*, for example, presented evidence for IT alpha-band oscillations having a pacemaker role for V4 alpha-band oscillations. Other research suggests that the efficacy of FEF influence on posterior regions may depend, in part, on long-range synchrony between these regions in the beta band.

The “method of visual inspection”

As noted previously, there is a striking similarity between the dorsal frontoparietal regions activated in the delayed-saccade task of Srimal and Curtis (2008; *Figure 8.4*), and the endogenous control network of Corbetta and Shulman (2002; *Figure 8.6*). Indeed, there have been many demonstrations in the literature that tasks engaging these two categories of behavior activate overlapping regions of the brain. However, there’s only so far that one can get by, in effect, saying, *See? These two patterns of activity look the same*. The reason for this is that brain activity – whether considered at the level of a single neuron, a cortical column, a population of 100 or so neurons, a 3 mm³ fMRI voxel containing tens of thousands of neurons, or a whole brain fMRI volume comprising ~30,000 such voxels – is simply too complex to be interpretable by visual inspection. Thus, for example, one cannot compare the patterns of activity illustrated in *Figure 8.4* and *Figure 8.6* and conclude, with any certainty, that the brain is “doing the same thing” during the two tasks that produced these images. Why? First, one can find many, many examples of experiments in which these same regions are active, but subjects are not performing either a spatial attention task or an oculomotor task. The IPS, for example, also shows elevated activity during tests of mental arithmetic and numerosity, of learning sequences of finger movements, of “understanding the goals of others,” of perceptual decision making, and so on. Second, what if activity that you can’t see determines the function that IPS carries out? Perhaps there’s different “unseen” activity happening in *Figure 8.4* vs. in *Figure 8.6* (recall, from *Methodology Box 6.1*, the uncolored parts aren’t inert, their activity simply isn’t exceeding a statistical threshold). Simply put, there is no brain area whose activity is specific for any single cognitive function. (“Take that, *Thought Box 1.1* and *Methodology Box 1.1* ...”)

Another way to think about this is that a limitation of the “visual inspection” approach is that it is inherently univariate. That is, we look at the blob in IPS in *Figure 8.4* and can only watch it vary along one dimension: its signal intensity either increases or decreases. This can lead to thinking of it as one “activation” doing only one thing (*Tip 8.15*).

Tip 8.15

Now I know that “*hate* is a strong word”; I frequently remind my kids of this. But this author hates the term “an activation,” because it implies a physical (and physiological) reality for something that doesn’t actually exist. One can measure *the activation* of a system, or *activity in a region*, but *an activation* implies the existence of a thing, like a cloud or a shadow. They do exist on the figures that we use to illustrate the results of our studies, but not in any biological sense. “Blob” would be a better term, because it can be used interchangeably, but the inherent silliness of the word keeps us from overinterpreting the significance of “an activation.”

Finally, *These two patterns of activity look the same* isn’t really what we care about. To address the question posed in the title of this section, what one would really want to be able to ask is *Is the brain actually doing the same thing during the performance of these two tasks?* One approach that offers a way to get closer to being able to ask this question is to apply one of the relatively recently developed methods of multivariate pattern analysis (MVPA) to one’s neuroimaging data. After first introducing some of the principles of MVPA in *Methodology Box 8.2*, we’ll wrap up this chapter by considering how one study has used it to ask this fundamental question.

“Prioritized Maps of Space in Human Frontoparietal Cortex”

This is the title of a study from the laboratory of Clayton Curtis, the same group that produced the delayed-saccade data illustrated in *Figure 8.4*. In it, Jerde et al. (2012) first employed a method similar to the one illustrated in *Figure 6.9.A* to identify cortical regions supporting retinotopic maps of space (*Figure 8.12.A*). Within the resultant ROIs, which included the dorsal frontoparietal regions associated with oculomotor control and with the endogenous control of attention, they applied MVPA to data from three tasks: oculomotor delayed response

METHODOLOGY BOX

8.2 Multivariate pattern analysis (MVPA) of fMRI data

Let's consider the hypothetical example illustrated in *Figure 8.11*, in which there are two voxels in a reduced dataset, voxel *F* and voxel *P*. On 10 trials our subject attends to the left, and on 10 trials she attends to the right. The first analysis that we run is to “collapse across” all the voxels in our ROI. That is, to average their signal together, then compare the spatially averaged signal from the 10 rightward trials vs. the spatially averaged signal from the 10 leftward trials. The bar graph in panel A of *Figure 8.11* illustrates that one can detect no appreciable difference between the two conditions (*Tip 8.16*). If we consider the two voxels separately, i.e., we determine the average signal from voxel *F* separately for the 10 leftward trials and the rightward trials, and do the same for voxel *P*, we seem to do a little bit better: keeping in mind that these are made-up data, paired *t*-tests indicate that voxel *F* shows significantly greater activity for leftward than rightward trials, and voxel *P* the converse. Nonetheless, each voxel also shows robust activity for its “nonpreferred” direction.

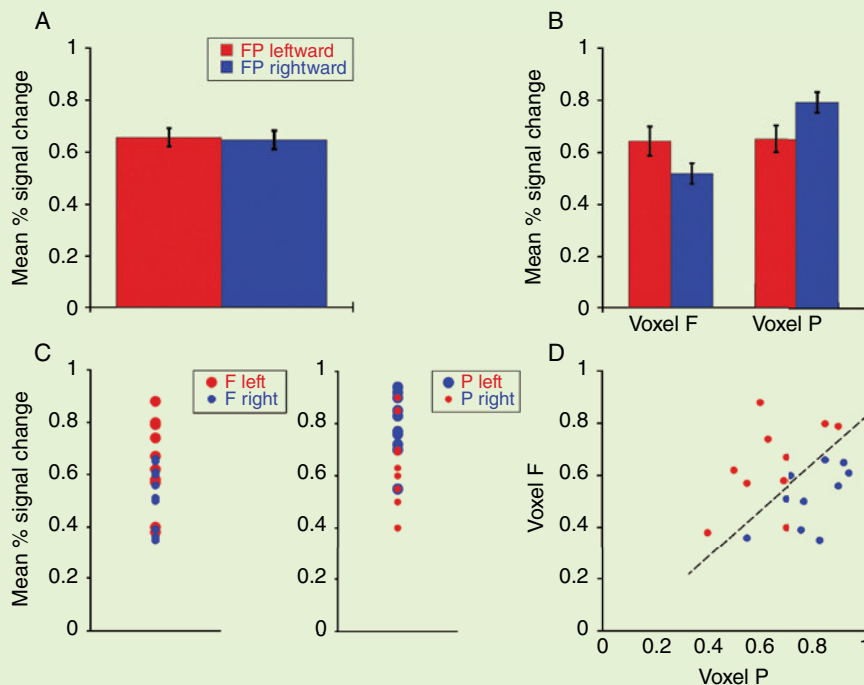


FIGURE 8.11 Improved sensitivity via MVPA. Plot **A** illustrates the data for the two conditions, averaged across voxel *F* and voxel *P*. In plot **B**, the data are broken out by voxel. The plots in **C** illustrate the signal intensity from each trial, for each voxel. Plot **D** is a multivariate re-representation of the data from **C**, with a discriminant line added to illustrate how one would classify a 21st trial, depending on which side of the line it falls. Source: © Bradley R. Postle.

Methodology Box 8.1 *Continued*

A consequence of this is that, if we were presented with data from a 21st trial, we'd have difficulty classifying it as a leftward or a rightward trial if we only knew that voxel *F*'s activity on this trial was, say, 0.6. To illustrate this, consider panel C, which shows each voxel's response on each of our 20 trials (these are the data that produced plots A and B): where, on the voxel *F* plot, would you draw the line above which you'd be confident that a trial was a leftward trial?

The two “stacks” of data points plotted in panel C are 1-D values – they only vary along a line. Panel D illustrates how the same data used to generate the plots in A, B, and C can be re-depicted as data that vary along two dimensions. To create the plot in Figure 8.11 D, all we've done is rotate the plot of voxel *P*'s activity – from panel C – from vertical to horizontal. That permits us to “project the data from each trial into a higher dimensional space.” (Sounds high-falutin', I know, but it *is* what we are doing; 2 is higher than 1 ...) Put more simply, it permits us to plot what each trial's values were at the two voxels simultaneously. As a result, it's easier to decide “where to draw the line”: if the hypothetical 21st trial falls above and to the left of the discriminant line that cuts through the cloud of data points, it's probably a leftward trial.

With real fMRI datasets, this same principle is applied, but the assertion that one is “projecting the data into a high-dimensional space” no longer elicits snickers, because, with, say, 300 voxels in an ROI, one is working in 300-D space. Within this space, **machine learning** algorithms find the high-dimensional hyperplane that, like our dotted line in Figure 8.11 D, provides the best discrimination between conditions A and B. The higher sensitivity afforded by MVPA will be exploited in many experiments that we will consider from here on out.

Tip 8.16

Note that this type of averaging across large numbers of voxels has been a bread-and-butter approach to cognitive neuroimaging, from its earliest days up through the day that these words are being typed. In this book, it has been illustrated in, e.g., Figures 8.4 and 8.5; and 6.8 and 6.18.

(ODR; “intention”; see, e.g., Figure 8.4); spatial “attention” (see, e.g., Figure 8.7); and spatial delayed recognition (short-term “memory”; e.g., Figure 8.6). (That is, after acquiring the retinotopic mapping data in one scanning session, the researchers had their subjects return for follow-up sessions in which they performed these three experimental tasks.) To address the question of *Is the brain actually doing the same thing during the performance of these three tasks?*, they employed a clever variant of MVPA, and produced a remarkable result.

A common use of MVPA is as a tool for “decoding” brain activity. Consider *Methodology Box 8.1*. After plotting the results of 20 trials of leftward vs. rightward attention, we felt reasonably confident that we would be able to classify the identity of a 21st trial, based solely on the neural activity associated with this trial. That is, we felt reasonably confident that we could *decode* the neural activity so as to know what the subject was thinking about. (Well, at least, so as to predict whether the subject was attending to the left or to the right.) What Jerde et al. (2012) did with their data is a step beyond “just” predicting the identity of a 21st trial drawn from the same task that produced the first 20 trials. Rather, what they did was train a classifier to discriminate between *attend-left* and *attend-right* trials, then asked it to classify trials from the intention task. That is, they defined the high-dimensional space into which *attend-left* and *attend-right* trials fell, established the high-dimensional plane that best discriminated *attend-left* from *attend-right* trials, then introduced trials from the ODR task and tried to predict, based on where the ODR data fell

in this attention-defined high-dimensional space, whether these were *saccade-left* or *saccade-right* trials. The only way that so audacious an approach could work would be if, in fact, *the brain was actually doing the same thing during the performance of both of these tasks.*

And it did work. Figure 8.12.B. illustrates that in sPCS (i.e., the FEF), and IPS2 (i.e., superior IPS), a multivariate classifier trained on only *intention* trials could decode data from *attention* trials, and vice versa. This finding provides some of the strongest evidence to date that “oculomotor control” and “spatial attentional control” may be two verbal labels that we use in different behavioral contexts but that, in fact, engage the same underlying neural activity. (This study and its implications are taken up in more detail in *Web Clip 8.3.*)

OF LABELS AND MECHANISMS

As humans, we like (need?) to categorize things. I just traded in my son’s “hybrid” bicycle (with properties of a “street bike” and of a mountain bike) for a (expensive!) mountain bike. To a nonexpert, the two would seem to be only superficially different from one another, but to my 11-year-old expert, they’re as

different as are a minivan from a sports car. To him, and to the biking community, they belong in different categories. To our intuitions, the act of moving one’s eyes and of choosing to focus one’s attention on object A vs. object B may feel like two different kinds of behavior. Certainly, up until quite recently, psychology and neuroscience have treated these as two different domains. But what if, as the data that we reviewed in this chapter would suggest, it is the very same underlying neural systems and processes that produce these two putatively different kinds of behavior? Surely, at least on the neural level, they can no longer be assigned to different categories. This is one of the reasons why so much attention, and often debate, is focused on seemingly arcane questions such as *Are PPC neurons encoding intention or attention?*, or *Are Moore’s attention-like results attributable solely to the activation of sensory neurons in the FEF, or do movement neurons also contribute?* This tension between different levels of description will reoccur often during the remainder of our time together. Just one example is manifested in Figure 8.12, which impels us to include spatial STM in this discussion of neural mechanisms vs. psychological categories, even though we’re not “supposed” to discuss it in detail until Chapter 13 ...

FIGURE 8.12 Is the brain “doing the same thing” during saccade planning and attentional orienting?

Source: Jerde, Merriam, Riggall, Hedges, and Curtis, 2012. Reproduced with permission of the Society of Neuroscience.

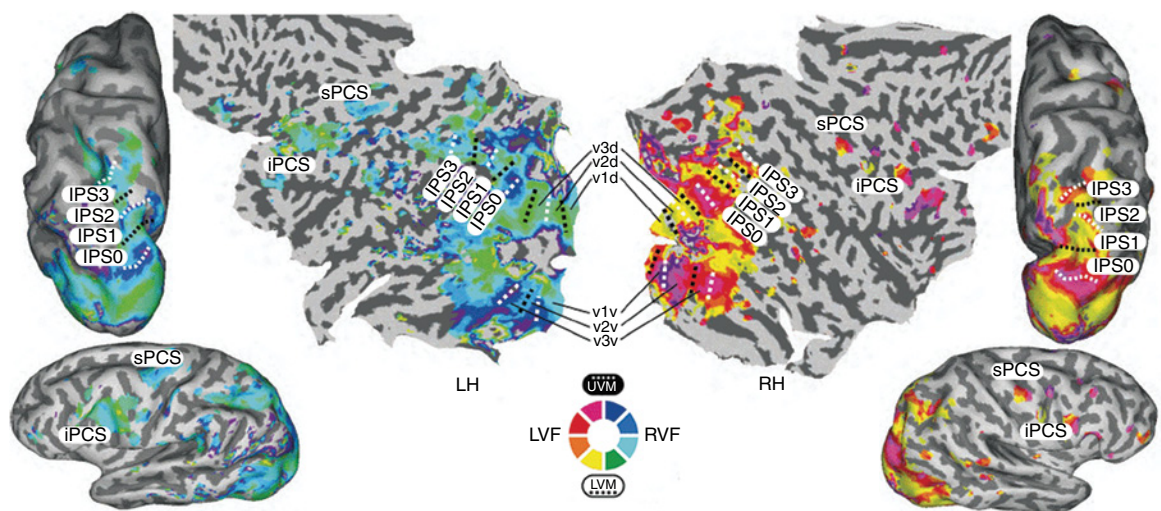


FIGURE 8.12.A The results of retinotopic mapping (compare with Figure 6.9), used to define the ROIs to which MVPA was applied to data from tests of “intention,” “attention,” and “memory.”

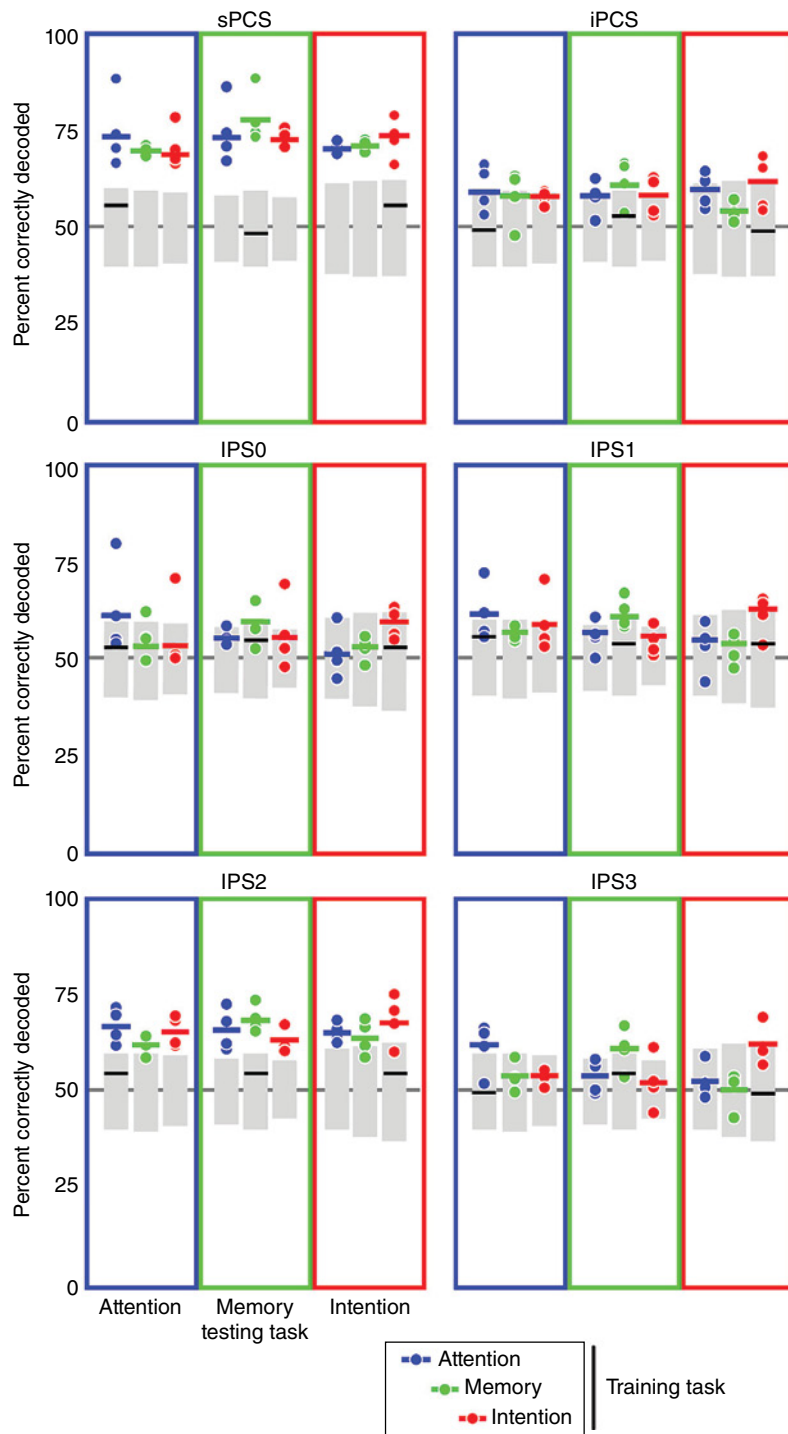


FIGURE 8.12.B Classification of data from tests of “intention,” “attention,” and “memory.” Each dot corresponds to a subject. The color of each set of dots indicates the task used to train a classifier; the color of each rectangle around a set of dots indicates the task used to test the classifier. Thus, e.g., the upper-left rectangle illustrates that, within sPCS, not only were *attention* trials decodable by the attention-trained classifier, but so, too, were *memory* trials and *intention* trials.

END-OF-CHAPTER QUESTIONS

1. What are some important differences between the posterior and anterior oculomotor control systems?
2. There are at least three different neural codes used to encode saccade metrics at different stations of the oculomotor system. Name the three codes, and the region(s) associated with each.
3. Define the exogenous vs. the endogenous control of attention. What cortical networks are associated with each?
4. How has the development of event-related methods for fMRI been critical for the study of oculomotor control and of attention? Describe one example from each domain.
5. Summarize the data that implicate the FEF as an important *source* of attentional control.
6. How does explicit computational modeling of the network interactions underlying visual search dispel the need to locate an attentional control homunculus in dorsal frontoparietal cortex?
7. How does MVPA differ fundamentally from neuroimaging data analysis methods that predominated during the decades of the 1990s and the 2000s? What advantages does MVPA offer?
8. Are “oculomotor control” and “attentional control” just two different labels that we have given to the same underlying neural mechanisms depending on the behavioral context in which they are used, or are there fundamental differences between the two? Defend your answer with data.

REFERENCES

- Chelazzi, Leonardo, John Duncan, Earl K. Miller, and Robert Desimone. 1998. “Responses of Neurons in Inferior Temporal Cortex During Memory-Guided Visual Search.” *Journal of Neurophysiology* 80 (6): 2918–2940.
- Corbetta, Maurizio, and Gordon L. Shulman. 2002. “Control of Goal-Directed and Stimulus-Driven Attention in the Brain.” *Nature Reviews Neuroscience* 3 (3): 201–215. doi: 10.1038/nnr755.
- Ferrier, David. 1875. “The Croonian Lecture: Experiments on the Brain of Monkeys (Second Series).” *Philosophical Transactions of the Royal Society of London*. 165: 433–488. doi: 10.1098/rstl.1875.0016.
- Ferrier, David. 1890. *The Croonian Lectures on Cerebral Localisation*. London: Smith, Elder and Co.
- Gandhi, Neeraj J., and Husam A. Katnani. 2011. “Motor Functions of the Superior Colliculus.” *Annual Review of Neuroscience* 34: 205–231. doi: 10.1146/annurev-neuro-061010-113728.
- Hamker, Fred H. 2005. “The Reentry Hypothesis: The Putative Interaction of the Frontal Eye Field, Ventrolateral Prefrontal Cortex, and Areas V4, IT for Attention and Eye Movement.” *Cerebral Cortex* 15 (4): 431–447. doi: 10.1093/cercor/bhh146.
- Ikkai Akiko, and Clayton E. Curtis. 2011. “Common Neural Mechanisms Supporting Spatial Working Memory.” *Neuropsychologia* 49 (6): 1428–1434. doi: 10.1016/j.neuropsychologia.2010.12.020.
- Jerde, Trenton A., Elisha P. Merriam, Adam C. Riggall, James H. Hedges, and Clayton E. Curtis. 2012. “Prioritized Maps of Space in Human Frontoparietal Cortex.” *Journal of Neuroscience* 32 (48): 17382–17390. doi: 10.1523/JNEUROSCI.3810-12.2012.
- Krauzlis, Richard J. 2005. “The Control of Voluntary Eye Movements: New Perspectives.” *Neuroscientist* 11 (2): 124–137. doi: 10.1177/1073858404271196.
- Moore, Tirin, and Katherine M. Armstrong. 2003. “Selective Gating of Visual Signals by Microstimulation of Frontal Cortex.” *Nature* 421 (6921): 370–373. doi: 10.1038/nature01341.
- Moore, Tirin, and Mazyar Fallah. 2001. “Control of Eye Movements and Spatial Attention.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (3): 1273–1276. doi: 10.1073/pnas.98.3.1273.
- Nabokov, Vladimir. 1955. *Lolita*. Paris: Olympia Press.
- Paré, Martin, and Douglas P. Munoz. 1996. “Saccadic Reaction Time in the Monkey: Advanced Preparation of Oculomotor Programs is Primarily Responsible for Express Saccade Occurrence.” *Journal of Neurophysiology* 76 (6): 3666–3681.
- Posner, Michael I. 1980. “Orienting of Attention.” *Quarterly Journal of Psychology* 32 (1): 3–25. doi: 10.1080/0033558008248231.
- Rizzolatti, Giacomo, and Laila Craighero. 2010. “Premotor Theory of Attention.” *Scholarpedia* 5 (1): 6311. doi: 10.4249/scholarpedia.6311.

- Schiller, Peter H. 1998. "The Neural Control of Visually Guided Eye Movements." In *Cognitive Neuroscience of Attention: A Developmental Perspective*, edited by John E. Richards, 3–50. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sherman, S. Murray, and Ray W. Guillery. 2006. *Exploring the Thalamus and Its Role in Cortical Function*. Cambridge, MA: MIT Press.
- Strimling, Riju, and Clayton E. Curtis. 2008. "Persistent Neural Activity During the Maintenance of Spatial Position in Working Memory." *NeuroImage* 39 (1): 455–468. doi: 10.1016/j.neuroimage.2007.08.040.
- Zarahn, Eric, Geoffrey K. Aguirre, and Mark D'Esposito. 1997. "A Trial-Based Experimental Design for fMRI." *NeuroImage* 6 (2): 122–138. doi: 10.1006/nimg.1997.0279.
- Zarahn, Eric, Geoffrey K. Aguirre, and Mark D'Esposito. 1999. "Temporal Isolation of the Neural Correlates of Spatial Mnemonic Processing with fMRI." *Cognitive Brain Research* 7 (3): 255–268. doi: 10.1016/S0926-6410(98)00029-9.

OTHER SOURCES USED

- Awh, Edward, Katherine M. Armstrong, and Tirin Moore. 2006. "Visual and Oculomotor Selection: Links, Causes and Implications for Spatial Attention." *Trends in Cognitive Sciences* 10 (3): 124–130. doi: 10.1016/j.tics.2006.01.001.
- Bastos, Andre, Conrado A. Bosman, Jan-Mathijs Schoffelen, Robert Oostenveld, and Pascal Fries. 2010. "Interareal Directed Interactions and Their Modulation by Selective Attention Assessed with High Density Electrocorticography in Monkey." Presented at the annual meeting for the Society for Neuroscience, San Diego, California, November 13–17.
- Bollimunta, Anil, Yonghong Chen, Charles E. Schroeder, and Mingzhou Ding. 2008. "Neuronal Mechanisms of Cortical Alpha Oscillations in Awake-Behaving Macaques." *Journal of Neuroscience* 28 (40): 9976–9988. doi: 10.1523/JNEUROSCI.2699-08.2008.
- Curtis, Clayton E., and Jason D. Connolly. 2008. "Saccade Preparation Signals in the Human Frontal and Parietal Cortices." *Journal of Neurophysiology* 99 (1): 133–145.
- Fischer, Burkhardt, and R. Boch. 1983. "Saccadic Eye Movements after Extremely Short Reaction Times in the Monkey." *Brain Research* 260 (1): 21–26. doi: 10.1016/0006-8993(83)90760-6.
- Funahashi, Shintaro, Charles J. Bruce, and Patricia S. Goldman-Rakic. 1989. "Mnemonic Coding of Visual Space in the Monkey's Dorsolateral Prefrontal Cortex." *Journal of Neurophysiology* 61 (2): 331–349.
- Hamilton, Antonia F. de C., and Scott T. Grafton. 2006. "Goal Representation in Human Anterior Intraparietal Sulcus." *Journal of Neuroscience* 26 (4): 1133–1137. doi: 10.1523/JNEUROSCI.4551-05.2006.
- Hamm, Jordan P., Kara A. Dyckman, Lauren E. Ethridge, Jennifer E. McDowell, and Brett A. Clementz. 2010. "Preparatory Activations across a Distributed Cortical Network Determine Production of Express Saccades in Humans." *Journal of Neuroscience* 30 (21): 7350–7357. doi: 10.1523/JNEUROSCI.0785-10.2010.
- Ikkai, Akiko, and Clayton E. Curtis. 2011. "Common Neural Mechanisms Supporting Spatial Working Memory." *Neuropsychologia* 49 (6): 1428–1434. doi: 10.1016/j.neuropsychologia.2010.12.020.
- Javal, Louis Émile. 1878. "Essai sur la physiologie de la lecture." *Annales d'oculistique* 80: 61–73.
- Kriegeskorte, Nikolaus, Rainer Goebel, and Peter Bandettini. 2006. "Information-Based Functional Brain Mapping." *Proceedings of the National Academy of Sciences of the United States of America* 103 (10): 3863–3868. doi: 10.1073/pnas.0600244103.
- Moore, Tirin, and Mazhar Fallah. 2004. "Microstimulation of the Frontal Eye Field and Its Effects on Covert Spatial Attention." *Journal of Neurophysiology* 91 (1): 152–162. doi: 10.1152/jn.00741.
- Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby. 2006. "Beyond Mind-Reading: Multi-Voxel Pattern Analysis of fMRI Data." *Trends in Cognitive Sciences* 10 (9): 424–430. doi: 10.1016/j.tics.2006.07.005.
- Pereira, Francisco, Tom Mitchell, and Matthew Botvinick. 2009. "Machine Learning Classifiers and fMRI: A Tutorial Overview." *NeuroImage* 45 (1–1): S199–S209. doi: 10.1016/j.neuroimage.2008.11.007.
- Poldrack, Russell A. 2006. "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10 (2): 59–63. doi: 10.1016/j.tics.2005.12.004.
- Rizzolatti, Giacomo, Lucia Riggio, Isabella Drascola, and Carlo Umiltà. 1987. "Reorienting Attention across the Horizontal and Vertical Meridians: Evidence in Favor of a Premotor Theory of Attention." *Neuropsychologia* 25 (1–1): 31–40. doi: 10.1016/0028-3932(87)90041-8.
- Shipp, Stewart. 2004. "The Brain Circuitry of Attention." *Trends in Cognitive Sciences* 8 (5): 223–230. doi: 10.1016/j.tics.2004.03.004.
- Squire, Ryan Fox, Nicholas A. Steinmetz, and Tirin Moore. 2012. "Frontal Eye Field." *Scholarpedia* 7 (10): 5341. doi: 10.4249/scholarpedia.5341.

FURTHER READING

- Aguirre, G. K., and M D'Esposito. 1999. "Experimental Design for Brain fMRI." In *Functional MRI*, edited by C. T. W. Moonen and P. A. Bandettini, 369–380. Berlin: Springer Verlag.
A comprehensive and accessible treatment of how and why the BOLD fMRI signal imposes constraints on experimental design, and of various design strategies that take these constraints into account.
- Awh, Edward, Katherine M. Armstrong, and Tirin Moore. 2006. "Visual and Oculomotor Selection: Links, Causes and Implications for Spatial Attention." *Trends in Cognitive Sciences* 10 (3): 124–130. doi: 10.1016/j.tics.2006.01.001.
An accessible, concise review of the cognitive neuroscience of attention with an emphasis on its relation to oculomotor control.
- Luo, Liqun, Eugenio Rodriguez, Karim Jerbi, Jean-Philippe Lachaux, Jacques Martinerie, Maurizio Corbetta, ... A. D. (Bud) Craig. 2010. "Ten Years of Nature Reviews Neuroscience: Insights from the Highly Cited." *Nature Reviews Neuroscience* 11 (10): 718–726. doi: 10.1038/nrn2912.
A "10 years on" commentary about subsequent developments in the two-attentional-pathways model from Corbetta and Shulman, whose 2002 review introducing this idea was the most highly cited paper to appear in Nature Reviews Neuroscience that year.
- Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby. 2006. "Beyond Mind-Reading: Multi-Voxel Pattern Analysis of fMRI Data." *Trends in Cognitive Sciences* 10 (9): 424–430. doi: 10.1016/j.tics.2006.07.005.
An accessible review of the logic underlying many implementations of MVPA and some applications to cognitive neuroscience questions.
- Pereira, Francisco, Tom Mitchell, and Matthew Botvinick. 2009. "Machine Learning Classifiers and fMRI: A Tutorial Overview." *NeuroImage* 45 (1–1): S199–S209. doi: 10.1016/j.neuroimage.2008.11.007.
An authoritative presentation of the principles of statistical machine learning that underlie many implementations of MVPA.
- Poldrack, Russell A. 2006. "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10 (2): 59–63. doi: 10.1016/j.tics.2005.12.004.
A prescient and thorough consideration of the logic of "reverse inference" that underlies the interpretation of functional neuroimaging data, and thereby provides a useful context from which to consider the relative advantages and limitations of univariate vs. multivariate approaches to neuroimaging data analysis. Although explicitly addressing neuroimaging, much of its argumentation is equally applicable to other types of neurophysiological data.
- Rizzolatti, Giacomo, and Laila Craighero. 2010. "Premotor Theory of Attention." *Scholarpedia* 5 (1): 6311. doi: 10.4249/scholarpedia.6311.
A "23 years on" commentary about subsequent developments relevant to this theory, including the role of skeletomotor systems.
- Sherman, S. Murray, and Ray W. Guillery. 2006. *Exploring the Thalamus and Its Role in Cortical Function*. Cambridge, MA: MIT Press.
Provocative text by two of the foremost authorities on the thalamus, proposing, among other ideas, that cortico-thalamo-cortical loops are the primary circuits for information transmission, with many cortico-cortical connections serving modulatory roles.

SECTION III

MENTAL

REPRESENTATION

INTRODUCTION TO SECTION III MENTAL REPRESENTATION

In *Section II* we covered “the inputs and the outputs.” And as I’ve stated previously, some would contend that everything from here on out is an elaboration of the principles that we’ve already learned. Well ... yes and no. One of the central features of the brain, its plasticity, means that it’s constantly changing, constantly updating its models of the world. A consequence of this is that no two instances of perceiving a stimulus are the same, because the system doing the perceiving is always somewhat different, somewhat changed, from the previous instance. At another level, *Section III* will examine ways in which perception and action are necessarily dependent on interpretations of the *meaning* of the inputs and the outputs. This idea was summarized eloquently in a commentary by the Catalán-born, US-based neurophysiologist Joaquín Fuster, as follows:

Perception is ... based on long-term memory. Every percept of the world around us is an “interpretation” of current sensory data in the light of our experience with similar data stored in long-term memory. Thus, perception has been appropriately called the testing of hypotheses about the world. We not only remember

what we perceive, but also perceive what we remember. Perceiving must essentially involve the matching of external gestalts to internalized representations in long-term memory. There cannot be an entirely new percept, because any conceivable new sensory configuration can resonate by associations of similarity or contiguity with others in long-term storage, which the new experience modifies and updates. In other words, the new experience modifies and updates previously established cortical networks. Intelligence and language, in all of their manifestations, operate in basically the same way, that is, by the activation of long-term memory networks towards the processing of intelligent behavior or speech. (Fuster, 2003, p. 737)

Long-term memory (LTM), the neural record of past experience that can influence future behavior, is the major focus of this section – how it’s formed; how it’s organized; how it influences cognition, perception, and action. Indeed, as an entrée into this topic, *Chapter 9* will, in effect, pick-up where *Chapter 5* left off, considering how the sensation and perception of a visual object interact with preexisting representations of the category, and perhaps the particular token, of the object being perceived.

REFERENCE

- Fuster, Joaquín M. 2003. "More Than Working Memory Rides on Long-Term Memory." *Behavioral and Brain Sciences* 26 (6): 737–737. doi:10.1017/S0140525X03300160.

CHAPTER 9

VISUAL OBJECT RECOGNITION AND KNOWLEDGE

KEY THEMES

- The neurological syndrome of visual agnosia raises many puzzling, and profound, questions about the relationship between perception and knowledge representation.
- Perhaps most intriguing is the seemingly selective inability of prosopagnosic patients to discriminate among individual faces.
- Despite the insight that can be gained from careful assessment of neurological deficits, models intended to explain “what must be broken” in these patients can sometimes posit cognitive mechanisms of questionable neural validity.
- Parallel distributed processing (PDP) models suggest that meaning (and, therefore, recognition) may reside within circuits that mediate perception and interactions between perception and action systems.
- fMRI evidence for a fusiform face area (FFA) that responds more strongly to faces than to other categories of stimuli has launched several vigorous debates about the neural and cognitive organization of the ventral stream.
- The FFA may also be sensitive to perceptual expertise, by which account its response profile would be better explained in terms of nurture than nature.
- MVPA analyses indicate that face information is also represented outside the FFA, and that the FFA can represent information about non-face categories, thereby challenging strong claims of the specificity of this region for face processing.
- (Despite the previous point,) direct electrical stimulation of the FFA, guided by electrocorticography (ECoG), selectively distorts the phenomenology of perceiving faces, thereby confirming a causal role of the FFA in face perception.

CONTENTS

KEY THEMES

VISUAL AGNOSIA

Apperceptive agnosia

Associative agnosia

Prosopagnosia

COMPUTATIONAL MODELS OF VISUAL OBJECT RECOGNITION

Two neuropsychological traditions

*Cognitive neuropsychology-influenced
models*

*Critiques of cognitive neuropsychology-
influenced models*

*Parallel distributed processing (PDP)-based
models offer a neurally plausible alternative
for modeling cognitive processes and
architectures*

The cognitive neuroscience revolution in visual
cognition

CATEGORY SPECIFICITY IN THE VENTRAL STREAM?

Are faces special?

"The Guns of August"

*Questions raised by the Kanwisher et al.
(1997) findings*

Perceptual expertise

Greeble expertise

Evidence for a high degree of specificity for many
categories in ventral occipitotemporal cortex

Evidence for highly distributed category represen-
tation in ventral occipitotemporal cortex

*MVPA challenges models of modular category
specificity*

Demonstrating necessity

RECONCILING THE IRRECONCILABLE END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

Although, as I promised, we won't begin this chapter with an overview of the anatomy underlying the function of interest in this chapter, we will use the approach, from *Chapter 6*, of introducing and motivating our topic with a neurological syndrome.

VISUAL AGNOSIA

In *Chapter 5*, it was noted that “By the 1960s, decades of clinical observations in humans and lesion studies in experimental animals had established that the temporal lobes play a critical role in object recognition.” Particularly germane for our present interests are the visual agnosias. This is a term for which understanding the etymology is more than a historical curiosity. *Gnosis* is ancient Greek for “knowledge,” and the *a-* in front of it confers negation. Thus, visual agnosia is the absence of visual knowledge. There are two broad categories of visual agnosia, a cardinal feature of both being profoundly impaired visual perception despite relatively intact visual sensation. That is, these patients aren't blind. The symptoms that make them agnostic reflect damage to cortical networks that are downstream from V1.

Apperceptive agnosia

A case of Benson and Greenberg (1969), Mr. S., was selected by Farah (*Tip 9.1*) to introduce the syndrome in her influential (1990) book on the topic. After suffering accidental carbon monoxide (CO) poisoning, Mr. S. retained the ability to detect small differences in luminance, wavelength (i.e., color), and size. However, he was “unable to distinguish between two objects of the same luminance, wavelength, and area when the only difference between them was shape.” Further, although failing to identify a safety pin, he reported that it was “silver and shiny like a watch or a nail clipper.” Thus, some low-level functions of sensation and feature identification were spared, as was, indeed, an abstract understanding of the visual properties of real-world objects. Additionally, “Recent and remote memory, spontaneous speech, comprehension of spoken language and repetition were intact,” indicating the absence of a general, nonspecific cognitive decline. And, most intriguingly, although “unable to name objects, pictures of objects, body parts, letters,

Tip 9.1

Martha Farah has pursued a remarkably wide-ranging career in cognitive neuroscience. Upon completion of her formal training in the cognitive psychology of visual perception (PhD 1983), she made herself, through collaborations and voluminous scholarship, a leading authority in the cognitive neuropsychology of visual perception in the 1980s and 1990s. Influential research using computational (in particular, connectionist) modeling and neuroimaging cemented her standing as a leading cognitive neuroscientist by the late 1990s. Around that time, she executed a “pivot” in her career trajectory, and has now established herself as a leading thinker in neuroethics and (from her Web site) “problems at the interface of neuroscience and society.”

numbers, or geometrical figures on visual confrontation, ... he could readily identify and name objects from tactile, olfactory, or auditory cues” (Benson and Greenberg, 1969, pp. 83–85).

An interesting question to ask at this point (one that we won't try to answer right away) is whether this syndrome may arise from a disconnection between early visual processing and the knowledge of the world that's necessary to recognize objects. After all, the knowledge might still seem to be intact in this patient, in that he can access it through other sensory modalities. One caveat to accepting this interpretation, however, is that it assumes a property of the cognitive architecture that need not be true: it assumes that knowledge about the world is stored in an “amodal” format, one that is not tied to any one sensory modality. Could it not be the case that, at least to some degree, some long-term representations of knowledge remain tied to the sensory channel through which we first acquired this knowledge? In other words, might it not be the case that Mr. S.'s *visual* knowledge of objects was compromised, but his tactile knowledge, olfactory knowledge, and auditory knowledge remained intact?

Other interesting variants of apperceptive agnosia are manifested by patients who can recognize objects in optimal viewing conditions, but whose performance declines precipitously under more challenging

conditions. For example, a patient may be able to recognize, say, a bicycle in a black and white line drawing, but fail at the same task if a piece of string is dropped onto the picture, or a handful of pencil shavings is scattered across it (i.e., if noise is introduced to the scene). Another classic deficit is the inability to recognize an object when it is presented from a non-canonical viewpoint. Thus, the patient may be able to recognize a wheelbarrow from a side view, but not a rear view (in which the handles are “pointing at” the viewer, perpendicular to the picture plane). Such deficits clearly implicate impairment of perceptual processes that we considered in *Chapter 5*, including viewpoint-dependent vs. viewpoint-independent object encoding, the disambiguating functions of top-down feedback, and, more generally, the hierarchical nature of object recognition.

Associative agnosia

This syndrome was famously characterized by (German-born, US-based) neuropsychologist Hans-Lukas Teuber (1916–1977) as “appear[ing] in its purest form as a normal percept that has somehow been stripped of meaning” (Milner and Teuber, 1968, p. 293). *Figure 9.1* shows an iconic example from a patient who was able to execute highly precise copies of line drawings, yet was unable to recognize what the drawings represented. Examples such as this very strongly give the sense of a disconnection syndrome (i.e., of perception from knowledge), as considered previously. However, although this idea (and, indeed, Teuber’s quote) remains prominent in the literature and in textbooks, it should be noted that the perception of many of these patients is unlikely to be truly normal. For example, despite the high fidelity of the drawings in *Figure 9.1*, many patients producing copies of comparable quality have been described as doing so in a “slavish,” “line-by-line” manner, and doing so “exceedingly slowly” (various authors, quoted by Farah, 1990, p. 60). Thus, the quality of the reproductions may owe more to “drawing strategies” than to truly intact visual perception.

It may be that the fundamental distinction between the two groups of patients (i.e., apperceptive vs. associative agnosics) isn’t that visual perception is impaired in one and spared in the other, but that they differ with regard to where along the visual processing

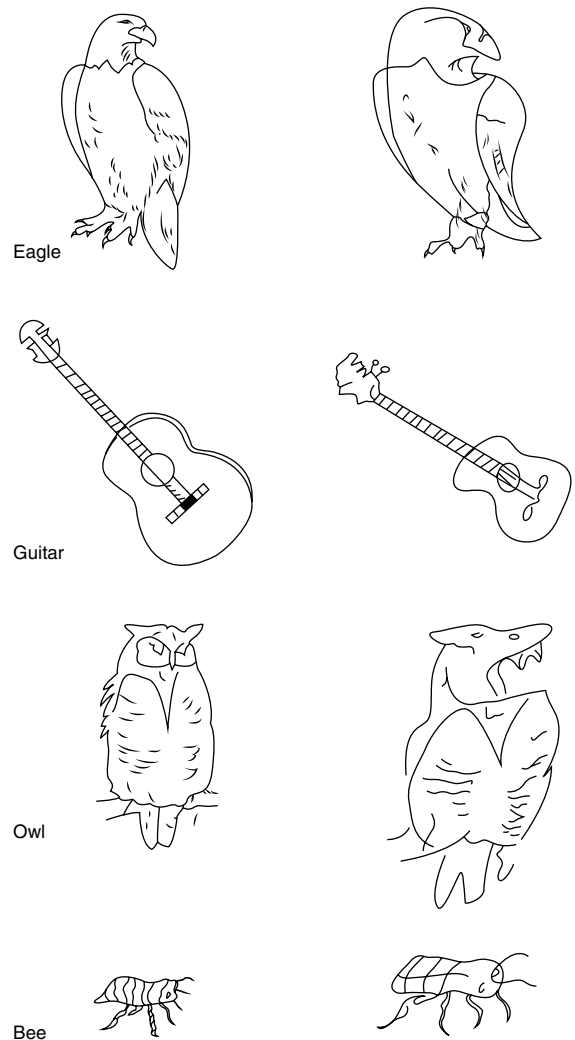


FIGURE 9.1 Drawings by a patient with an associative agnosia. On the left are the models and on the right are the patient’s copies. Despite reasonably good reproduction of all the visual features, the patient was unable to name the objects that they represent. He named the eagle “a cat sitting up,” the guitar “some kind of machine, a press,” the owl “a pattern,” and the bee “an animal with horns and a tail ... a rhino?”

hierarchy their impairment is most pronounced. For example, whereas apperceptive agnosic patients try to identify objects based on low-level features (e.g., color, texture) associative agnosic patients typically try to rely on shape. Thus, for example, Farah (1990)

notes that on four separate occasions she showed the associative agnosic patient L. H. a drawing of a baseball bat, and on each his incorrect guess was of an object with a similar shape: a paddle, a knife, a baster, and a thermometer. *Web Link 9.1* features a neuropsychological exam with an associative agnosia patient.

The most intriguing subset of associative agnosic patients are those whose primary deficit seems not to be with type identification, but with token individuation. And within this subset, the sub-subset whose impairment very saliently includes face recognition has claimed a disproportionate share of the attention of the visual cognition community.

Prosopagnosia

Following in the tradition of the German neurologist Heinrich Lissauer (1861–1891), who coined the term agnosia, it was also a German neurologist, Joachim Bodamer who, in 1947, welded on the prefix “proso-pon” (from the ancient Greek for face) to characterize his patients with a selective agnosia for the recognition of faces (*Tip 9.2*). In the decades that followed, there have been many case studies of patients with this syndrome. In many cases the deficit is quite striking, with the patient, for example, not recognizing his wife as she walks past the hospital bed, and not even realizing that the strange person staring at him is actually a reflection of himself in a mirror. *Web Link 9.2* presents a compelling profile of one such patient, managing to be simultaneously very insightful and very poignant.

What over a half-century of research hasn’t resolved, however, is just how specific to faces this syndrome can be. There is, for example, a report of a patient who could identify close relatives by voice, but not by face, but who had no difficulty recognizing, for example, his electric razor from among many similar-looking electric razors, his wallet from among similar wallets, and so on. Intriguing dissociations between patients have also been reported, such as one group with a face-recognition deficit but who have no difficulty recognizing facial expressions, and others showing the

converse. Many other reports, however, indicate that within-category recognition of other categories can also be impaired, with animals and plants often compromised. Specific examples include an avid bird-watcher who, upon developing prosopagnosia, also lost the ability to discriminate between different species of birds. Perhaps most famous is the case of the farmer who, along with human faces, lost the ability to recognize each of the cows in his herd. This and other reports gave rise to the idea that prosopagnosia and other seemingly category-specific deficits of visual perception may reflect a more general loss of perceptual expertise, an idea that we’ll take up further along in this chapter.

COMPUTATIONAL MODELS OF VISUAL OBJECT RECOGNITION

Two neuropsychological traditions

Interestingly, although much of the research summarized in the preceding sections occurred in parallel to the neurophysiological study of object recognition in the monkey, including studies of face processing (*Chapter 5*), the two traditions of research proceeded largely independent of one another. One reason is that they had largely different goals. Whereas systems neuroscience, including the subfield that I’ll call experimental neuropsychology, prioritizes understanding how the brain works, the discipline of cognitive neuropsychology is primarily interested in how the mind works (i.e., in how the cognitive system is organized). Examples of experimental neuropsychology that we have already considered include the studies cited in *Chapter 5* as influences of Pohl (1973), and the research by Ramachandran on patients with phantom limb syndrome (*Chapter 4; Tip 9.3*). In relation to visual agnosia, see Milner and Teuber (1968). An example of cognitive neuropsychology, in contrast, is the influential model of object recognition from Humphreys and Riddoch (1987), depicted in *Figure 9.2.A. (Tip 9.4.)*.

Tip 9.2

As has been the case for earlier generations of neurologists, Bodamer’s patients were veterans who had suffered bullet wounds to the head.

Cognitive neuropsychology-influenced models

Note from *Figure 9.2.A* that the emphasis of this model is on stages of information processing, rather than stages of neural processing. From this perspective, although it is intended to account for object recognition

deficits reported from a variety of brain-damaged patients, it is addressing a level of explanation that is more closely related to cognitive psychology than to neuroscience. Thus, for example, each of the boxes in this model are inferred, rather than directly measured.

FIGURE 9.2 Computational models of object recognition.

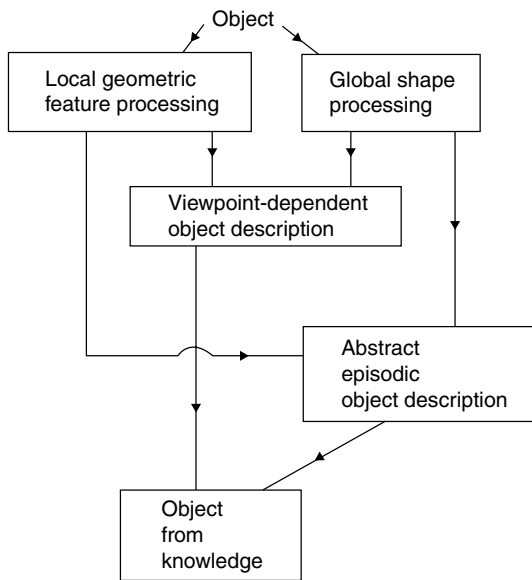


FIGURE 9.2.A Cognitive neuropsychology-derived model of visual object processing. Source: Humphreys and Riddoch, 1987. Reproduced with permission of Lawrence Erlbaum Associates.

They correspond to hypothesized stages of processing that are proposed in order to account for one or more findings from visual agnostic patients.

To summarize, Humphreys and Riddoch (1987) offer an analogy that's evidently a popular one – searching for a book at the library in the pre-internet age. (Because this author graduated from university in 1987, it is an analogy that he understands quite well ...) When one wants to find a particular book at the library, one looks it up by title or by author in the card catalog, and the card corresponding to that book indicates the specific location in the stacks (the “address”) where the book will be located. (In the US, this

Tip 9.3

A nice example of an experimental neuropsychology approach to visual agnosia is Teuber's “Perception” section in Milner and Teuber (1968).

Tip 9.4

A third discipline/vocation sharing this word is *clinical neuropsychology*, which entails the administration of carefully designed tests (often referred to as “neuropsych testing”) to diagnose and quantify behavioral impairments in brain-damaged patients, and to track recovery and/or disease progression in patients.

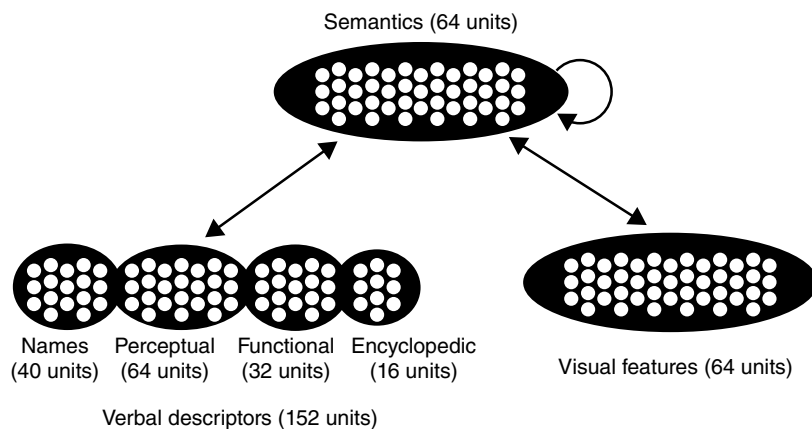


FIGURE 9.2.B PDP model of visual cognition. Source: Rogers, Lambon Ralph, Garrard, Bozeat, McClelland, Hodges, and Patterson, 2004. Reproduced with permission of APA.

address is encoded via the Dewey decimal system.) The library patron writes down the encoded address on a scrap of paper, then begins to search the stacks. In this analogy, the finding of the appropriate card in the card catalogue corresponds to the transduction of visual information into a neural code, and the writing of the address on to the scrap of paper is early visual processing – the first three boxes in the model. The scrap of paper containing the link to the book is the “abstract episodic object description,” which corresponds to a higher-level visual object representation. (In the analogy, this representation is *abstract* because the Dewey code doesn’t literally say “third stack from the north wall, fourth shelf, 437 books in from the end”; nor does it have any meaning other than as a guide to help you find your book – therefore, this level of perceptual representation is posited to be without inherent meaning.) The key step in the process is recognition, when one matches the code on one’s scrap of paper with the code on the spine of the sought-after book. (The code on the spine of the book corresponds to “object form knowledge” in the model.) Recognition having successfully occurred, the patron now has access to the meaning of the card in the catalogue – the contents of the book. Phenomenologically, this corresponds to recognizing the object that one is looking at, and knowing everything that there is to know about it. (E.g., *That’s my cat. Cats are quadrupedal domesticated mammals, probably bred from the lynx or ocelot or some smallish feline carnivore [etc.]. This cat’s name is Mahalia, named after Mahalia Jackson because she’s black, brown, and beige (the cat, that is), the title of a Duke Ellington piece (on which M. J. sang) that we were into when we “rescued” her as a kitten from that crazy cat lady’s house in Billerica, MA. She’s really old (again, referring to the cat) ...*) Associative agnosia, by this account, is caused by damage to object form knowledge. (Note that this is the interface that Fuster (2003), in the *Introduction to Section III*, refers to as “the matching of external gestalts to internalized representations in long-term memory” that is central to perception.)

Critiques of cognitive neuropsychology-influenced models

Farah (1990) notes that this type of modular, “symbol-manipulation” model (*Tip 9.5*) “has appealed to many [cognitive] neuropsychologists seeking an account of

Tip 9.5

Sometimes these are pejoratively referred to as “box-and-arrow” models.

associative agnosia” because it provides a framework for the intuitions of observers of the syndrome. She summarizes these, drawing on literature spanning from 1914 through 1985, as “loss of the ‘stored visual memories’ or ‘engrams’ of objects,” and disruption of “‘the matching of signal to trace’ ... [and the] destruction of stored ‘templates’” (pp. 96–97). However, there are at least two types of concern that one could have about such models. (And this is now me talking again, not Martha Farah). One is that they are not quantitative, and thus are difficult to disprove. With all models, and certainly with “verbal models” such as this, one must guard against post hoc elements that satisfy one’s intuition about how a system might work. For our current discussion, make note of the fact that an explicit tenet of the Humphreys and Riddoch (1987) model, embodied in the *abstract episodic object description*, is that visual object recognition proceeds in two stages: (1) the processing of “structural” representations that have no inherent meaning, and (2) “visual semantics,” the portion of long-term memory that represents meaning. A second concern about modular models is that they are often not constrained by/do not incorporate what we know about how the brain works. The model in *Figure 9.2*, for example, has no feedback from higher levels of visual processing to lower levels. Further, it contains no specification that a certain kind of neuron or area of the brain (such as those we reviewed in *Chapter 5*) corresponds to any of the processing stages in the model.

Parallel distributed processing (PDP)-based models offer a neurally plausible alternative for modeling cognitive processes and architectures

At this juncture, a fair objection might be to remind me that the goal of a model such as that illustrated in *Figure 9.2* is to better understand behavior at a *cognitive* level of description, not at a *neural* level. And it is certainly true for such models that slapping “neuro-sounding” labels onto a model won’t, in and of itself, confer any more legitimacy or plausibility to it.

METHODOLOGY BOX

9.1 Parallel distributed processing (PDP) modeling of cognition

As illustrated in *Figure 9.2.B*, PDP models are made up of groups of interconnected processing units that stand in for neurons in that they achieve levels of activity determined by the inputs that they receive. (Depending on the model, all units can be “excitatory,” or “inhibitory” units may be included.) These units are nonlinear, like neurons, in that their activation level is determined by some variant of a sigmoid function. (I.e., if the activation state is really low or really high, one or two additional inputs won’t make much difference; the midrange is the most sensitive.) A critical feature of PDP models is that the strength of connections between units can change as a function of, e.g., how effective unit A is at influencing the activity level of unit B. This captures a fundamental property of real neural connections (as we’ll see in *Chapter 10*), and endows these models with a very important property – they learn.

The architecture of the PDP model in *Figure 9.2.B*, with a visual layer, a semantic layer, and a verbal layer, is based on the authors’ interpretation of the proposal of the German neurologist Karl Wernicke in 1900 (as cited in Eggert, 1977):

The concept of a rose is composed of a “tactile memory image” ... in the ... somesthetic cortex ... [and] of a visual memory image located in the visual ... cortex. The ... repetition of similar sensory impressions results in ... a firm association between those different memory images [such] that the mere stimulation of one sensory avenue by means of the object is adequate to call up the concept of the object ... This sum total of closely associated memory images must “be aroused into consciousness” for perception not merely of sounds of the corresponding words but also for comprehension of their meaning. (p. 237)

Thus, in the model in *Figure 9.2.B*, seeing a particular visual stimulus is simulated by activating a unique pattern of units in the “visual features” layer, and hearing a particular word is simulated by activating a unique pattern of units in the “verbal descriptors” layer. Because the semantic units do not receive direct, external inputs from the environment, they are referred to as hidden units. The **hidden unit layer** is another fundamental feature of PDP models, and is, if you will, “where the magic happens.” Training is implemented by presenting the model with a visual (or verbal) stimulus, evaluating how accurately (after propagating signals through the network) it produces a verbal (or visual) output, and, when it makes an error, delivering a “teaching signal” that adjusts the weights of connections within the model (strengthening those that contributed to “correctness” in the response, and weakening those that contributed to error in the response). (Such iterative adjusting of weights to fine-tune performance is reminiscent of the computations implemented by the cerebellum.) The end result is that individual “meanings” as defined by Wernike (1900; cited in Eggert, 1977) don’t exist in any one place. Rather, they are distributed across the patterns of weighted connections throughout the network. (Note how the issues highlighted here hearken back to the tension introduced in *Chapter 1* between localization of function vs. mass action; *Tip 9.6*.) See *Web Clip 9.1* for a step-by-step demonstration of how a PDP model can be trained to recognize objects.

RESEARCH SPOTLIGHT

9.1 Where's the recognition in visual object recognition?

Rogers, Hodges, Lambon Ralph, and Patterson (2003) note that many neuropsychologically inspired models of object recognition posit that:

a stimulus object or word must be recognized before its meaning can be accessed. "Recognition" in turn is construed as the process of matching a perceptual representation of the stimulus item to stored representations of previously encountered stimuli (or representations derived from these). Such stored "structural" representations are understood to be qualitatively different from the "semantic" representations that store meanings – they do not encode explicit semantic content, but capture information about visuo-spatial structure. (p. 626)

The model illustrated in *Figure 9.2.A* is an example. The PDP model of Rogers et al. (2004; *Figure 9.2.B*), in contrast, tests an alternative: "semantic knowledge does not reside in representations that are separate from those that subserve perception and recognition, but emerges from the learned associations amongst such representations in different modalities. Neuroanatomically, the semantic system ... consists in those cortical tracts and regions that permit the interaction of perceptual representations in different modalities" (Rogers et al., 2003, p. 626). From *Chapter 5*, we know that candidate regions for such interactions include IT, and, certainly, STP.

The model was trained with 48 simulated items, half corresponding to living things (birds, mammals, fruits) and half to nonliving (vehicles, household objects, tools). Each item was presented both as a visual object and as a verbal label once per block, for 400 blocks. Upon completion of training, the model could be presented with any stimulus in either modality, and produce the appropriate response in the other modality. E.g., present the word "apple" to its verbal layer, it produces the representation of *apple* in its visual layer, and vice versa. The key implication of this result is that the model achieved object recognition without explicitly containing two kinds of "internal" representation – those that encode structural information sufficient to support stimulus recognition, and those that encode semantic content, or "meanings." This challenges not only explicit neuropsychological models (e.g., *Figure 9.2.A*), but also the common intuition held by scientists and nonscientists alike, conveyed in the Fuster (2003) statement, that "Perceiving must essentially involve the matching of external gestalts to internalized representations in long-term memory." The need for putative "grandmother" units is obviated.

After achieving this result in a "healthy" network, Rogers and colleagues (2004) then simulated neuropsychological findings by "lesioning" the model after it had been trained, thereby simulating brain damage acquired in adulthood, after one's long-term memory for common objects had been fully formed. The primary clinical interest of this study was a condition variously known as *semantic dementia* or *primary progressive aphasia*, a neurodegenerative disorder characterized by broad cortical atrophy (more detail in *Chapter 12*). Simulated lesions were produced by randomly removing varying proportions of connections between units at all levels in the model. (To simulate visual agnosia, one might restrict the "lesioning" to only connections between the "visual features" layer and the hidden layer; e.g., Plaut, 2002.) In subsequent simulations of neuropsychological tests – naming, category sorting, word-to-picture matching, and drawing from memory – the lesioned PDP model showed patterns of deficits that were qualitatively similar to the performance of patients with semantic dementia/primary progressive aphasia.

Tip 9.6

Although rarely as extreme as Florens in denying any localization of function, contemporary PDP modelers do often challenge assumptions of symbolic, modular architectures (see *Research Spotlight 9.1*).

However, *Methodology Box 9.1* and *Research Spotlight 9.1* will introduce an alternative type of computational model – the parallel distributed processing (PDP), or connectionist, model – that, although also often addressing questions at the cognitive level of description, are generally considered to be more compatible with many of the assumptions and practices of contemporary cognitive neuroscience. And, as we'll see in *Research Spotlight 9.1*, a substantive consequence of employing a PDP model is that it will draw very different conclusions about how the “matching of external gestalts to internalized representations in long-term memory” happens in visual object recognition. (To be fair to Humphreys and Riddoch, more recent incarnations of their models now also include more neuroscientifically informed constraints. The didactic goal here is to emphasize how the choice of modeling framework can strongly influence, well, one's model of how a particular cognitive process is organized.)

For this author, the really powerful implication of *Research Spotlight 9.1* is the way that it encourages a radical rethinking of how object recognition works. While writing *Chapter 5*, for example, I was feeling impelled to press “forward” (i.e., downstream) to “get to the place where recognition happens.” Unbeknownst to me, as it were, I may have been describing *where*, if not *how*, object recognition happens while detailing feedforward and feedback connections from V1 to V4 and MT, from V4 to IT, and the response profiles of neurons all along these pathways. There may, in fact, have been no finish line to cross ...

The cognitive neuroscience revolution in visual cognition

With these ideas in mind, we now return to the physiology of the ventral pathway, this time (as opposed to in *Chapter 5*) as measured with fMRI, and in the

human. This next section will illustrate the truly revolutionary influence that developments in cognitive neuroscience, in this case neuroimaging, have had on fields such as object recognition and visual cognition. One example is that the divide between the cognitive neuropsychology and the neurophysiology of object recognition, which we have just reviewed, has been steadily narrowing. Indeed, it is increasingly rare today to see models of object recognition and/or visual agnosia that don't explicitly incorporate constraints from functional neuroanatomy and/or neurophysiology. In other respects, however, some longstanding debates persist, just reframed in terms of the newer technologies and the newer intellectual paradigms. Thus, for example, questions of localization-of-function vs. distributed processing remain front-and-center, and the even older dialogue between nature and nurture is also at play in this arena.

CATEGORY SPECIFICITY IN THE VENTRAL STREAM?

Are faces special?

“The Guns of August”

You'll recall, from *Chapter 5*, the puzzling properties of neurons of temporal cortex that seemed to respond to “the idea of a face” more so than to any individual face. In the 1990s, cognitive psychologists familiar with this literature, and with the puzzling questions raised by the prosopagnosic syndrome, saw in neuroimaging an opportunity to address these problems in a new way. Now I say “cognitive psychologists” – and Farah, in later additions of her 1990 book, chronicles many of the early contributors – but much of the intellectual and empirical energy driving these “early” cognitive neuroscience developments was one individual: Nancy Kanwisher. Kanwisher trained as a cognitive psychologist, and early in her career worked on a phenomenon known as repetition blindness, influentially relating it to how the brain handles the type vs. token distinction (e.g., why is it more difficult to individuate two different flowers presented sequentially in a serial stream of stimuli than, say, a flower and a face?) (Kanwisher 1991). Perhaps this underlies her subsequent fascination with face processing (*Tip 9.7*). In any event, in

Tip 9.7

Nancy Kanwisher embodies “living the courage of one’s scientific convictions.” In the early 1990s, convinced that neuroimaging offered a unique opportunity to study visual cognition, she left a tenured faculty position at a prestigious psychology department in the US for an untenured position at Harvard – which, at the time, was notorious for not promoting its untenured faculty, no matter how brilliant or accomplished – because Harvard was willing make scanning time available to cognitive psychologists. (Even so, in “the early years,” many of her group’s scans were performed in the wee hours of the morning, when users with higher priority didn’t want to be working.)

Tip 9.8

Perhaps for confidentiality reasons, and also to be clever, the faces portrayed in the figures are of then-student Ewa Wojciulik and co-authors McDermott and Chun.

1997 she, together with then-postdoctoral fellow Marvin Chun and then-undergraduate research assistant Josh McDermott (now each a highly accomplished cognitive neuroscientist in his own right) published the study featured in *Research Spotlight 9.2* that spurred a subfield of cognitive neuroscience that shows no sign of slowing down 15 years later.

One reason that the discovery of the FFA generated so much interest was that it seemed to effect a neural test of an influential theory that had been

RESEARCH SPOTLIGHT

9.2 “The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception”

The proclamation of this title of the Kanwisher, McDermott, and Chun (1997) study remains provocative and controversial to this day. The study employed a block design, in which subjects viewed the serial presentation of unfamiliar faces (ID photos of Harvard undergraduates; *Tip 9.8*) alternating with serial presentation of control stimuli. Blocks were 30 seconds long; stimuli presented for 500ms with an interstimulus interval of 150ms. The experiment proceeded in three phases, the first an unapologetic “far and wide” search for regions that might respond with greater activity for faces vs. objects, the second and third more rigorous tests to replicate the phase 1 finding while ruling out potential confounds and further assessing the specificity of the regions identified in phase 1. The authors note that “Our technique of running multiple tests applied to the same region defined functionally within individual subjects provides a solution to two common problems in functional imaging: (1) the requirement to correct for multiple statistical comparisons and (2) the inevitable ambiguity in the interpretation of any study in which only two or three conditions are compared” (p. 4302). These statements are unpacked in *Methodology Box 9.2*.

In phase 1 of the study, 12 of 15 subjects showed significantly greater activity for faces than objects in the right fusiform gyrus, on the ventral surface of the temporal lobe, at the level of the occipitotemporal junction, and this region therefore became the focus of the subsequent scans. Phase 2, illustrated in *Figure 9.3*, panels 1B and 1C, was intended to rule out the potential confounds that the phase 1 results were due to (1) differences in luminance between face and object stimuli, or (2) the fact that the face, but not the object, stimuli were all exemplars drawn from the same category. Phase 3, illustrated in *Figure 9.3*, panels

Research Spotlight 9.2 Continued

2B and 2C, was intended to rule out the potential confound that the FFA was simply responding to any human body parts. Additionally, it established that FFA activity generalized to images of faces that were less easy to discriminate, and to a condition that imposed stronger attentional demands on the processing of stimuli, by requiring a decision of whether or not each stimulus matched the identity of the previous one.

As illustrated in *Figure 9.3*, the results were quite striking, showing considerably higher activity in FFA to faces than to any other stimuli. Thus, they are consistent with the idea that the faces receive privileged processing in this region. But is this region *specialized* for face processing? Sometime in the late 1990s, this author attended a presentation by Kanwisher at a colloquium at the University of Pennsylvania. The cognitive scientist Saul Sternberg asked what to make of the fact that the region also shows elevated activity (relative to baseline) to other stimuli. Kanwisher replied (and here I'm paraphrasing from memory) that “there must be something to this question, because I get it almost every time I give this talk,” but that quantitative superiority of response to a particular class of stimuli has always been interpreted by physiologists as evidence for specificity of function. This stance will come under closer scrutiny in later portions of this chapter.

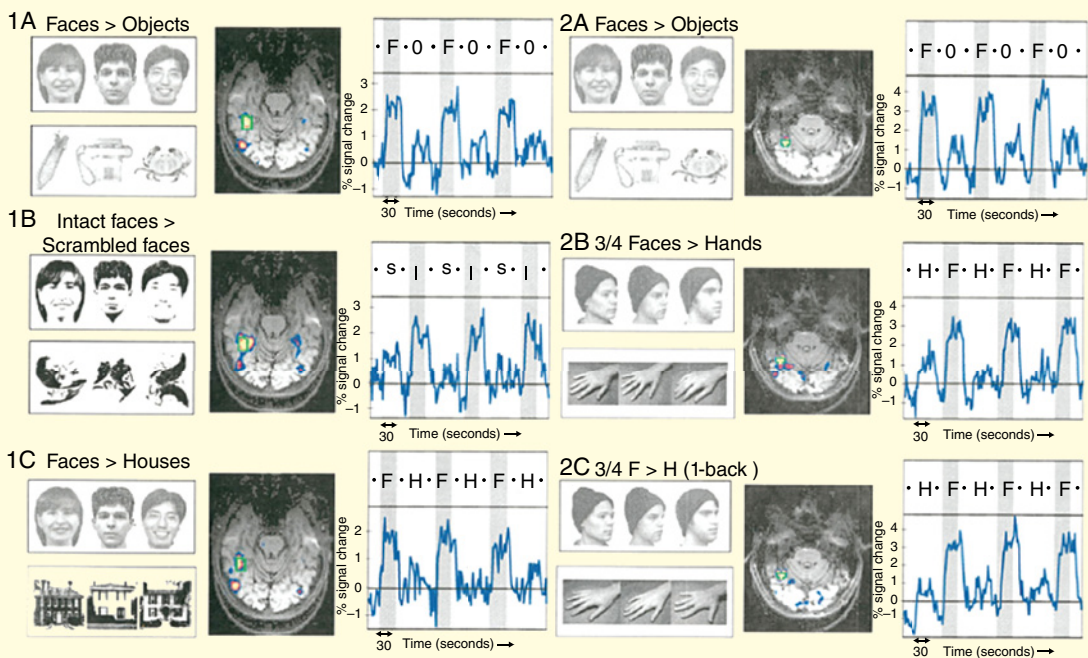


FIGURE 9.3 Examples of stimuli, fMRI statistical maps, and time series, from several assays of response properties of the FFA. The contrast illustrated in panels **1A** and **2A** (faces > objects) was used to define the FFA ROI, outlined in green in the axial images, each for a different individual subject. Panels **B** and **C** illustrate the statistical map generated by each of the contrasts, each for the same subject, with the green-outlined ROI from panel **A** superimposed. This procedure was repeated on the individual-subject data of five subjects, and each time series is the average from the FFA ROI of each. Source: Kanwisher, McDermott, and Chun, 1997. Reproduced with permission of the Society of Neuroscience.

METHODOLOGY BOX

9.2 Some problems, and solutions, with univariate analyses of fMRI data

Research Spotlight 9.2 raises two common problems encountered when designing an fMRI study. Problem (1) is easiest explained with reference to *Figure 8.11*. As alluded to in *Methodology Box 8.2*, the plots in *Figure 8.11 A* and *B* result from univariate analyses. In *Figure 8.11 B*, for each of the two voxels, signal intensity was statistically compared for leftward trials vs. rightward trials. With conventional hypothesis testing, this is done by assuming the null hypothesis that the signal in these two conditions doesn't differ, and rejecting the null hypothesis if the probability that their numerical difference doesn't differ is determined to be less than .05. That is, the scientific community has generally agreed to accept a procedure for which, once every 20 analyses, an effect will be determined to be "statistically significant" even though it is not. Now what if, instead of two voxels in *Figure 8.11 B*, there were 20? And further, what if (because I made up the values for these voxels myself) I told you that there is no difference between leftward vs. rightward trials in any of them? (They're my voxels, so I decide.) Were you, nonetheless (perhaps out of spite?), to go ahead and perform a *t*-test at every voxel, chances are that you'd get a significant result for one of them. (That's what $p = .05$ means.) Conventional univariate analyses of fMRI data sets are referred to as "massively univariate," because a separate statistical test (some variant of the general linear model (GLM), comparable to a *t*-test) is performed at every voxel. At the level of a whole brain, which typically has >20,000 voxels, one would expect that 1000 or more voxels would be falsely identified as being "activated" unless one were to apply a correction. The most rigorous is to divide each *p* value by the number of tests (called Bonferroni correction). For just two voxels, this means that a critical *p* of .025 needs to be achieved. Not a big deal. But for 20,000, the critical *p* would drop to .00000025. In *Figure 8.11 A*, in contrast, because the activity from the two voxels is averaged into one ROI, no correction needs to be applied, because only one test is being performed. Therefore, defining an ROI with one contrast, then "interrogating" it with either data from a different block, or a contrast that is mathematically orthogonal to the one used to define the ROI, is an effective and frequently used strategy.

Problem (2) simply refers to the fact that within a typical scanning session of 1 to 2 hours, there's a limited number of comparisons that one can make. To make a strong claim that a region is specific ("specialized") for processing faces, however, one wants to be able to show that face-evoked activity is greater than activity for many, many other classes of stimuli. The solution of Kanwisher, McDermott, and Chun (1997) was to bring the same subjects back for multiple scanning sessions, evaluating responses to different stimuli each time. (Note that, strictly speaking, to establish the absolute specificity of the FFA for faces, one would have to assess the region's response to every conceivable stimulus on the planet, as well as every conceivable brain state ...)

advanced by the philosopher Jerry Fodor in his 1983 book *Modularity of Mind*. In a sentence, Fodor's idea was that particular cognitive operations were "encapsulated" and "cognitively impenetrable" once launched, and thus not susceptible to influence by other factors (attention, intention, behavioral priorities, distraction,

...) until they were accomplished. Visual object recognition, as we've been emphasizing here, and auditory perception (particularly speech perception), would be two such candidate processes. Not unlike some of the cognitive neuropsychological models that we considered earlier in this chapter, Fodor's idea gave a good

Tip 9.9

Although Fodor's ideas about modularity are explicitly related to the cognitive architecture, there are certainly conceptual similarities between them and localizationist views of how the brain works. Indeed, the frontispiece of Fodor's 1983 book is a phrenological bust, thereby making explicit the association of its ideas with the scientific legacy of Gall.

account of many observations from studies of human behavior (in this case, from cognitive psychology), but also seemed contrary to much of what we know about how the brain works. Thus, one really intriguing aspect of the Kanwisher et al. (1997) study was that it could be interpreted as a neural confirmation of a hypothesis that seemed contrary to how many thought that the brain works! (*Tip 9.9.*)

Questions raised by the Kanwisher et al. (1997) findings

Now, to be clear, no one questioned the veracity of the Kanwisher, McDermott, and Chun (1997) findings. Rather, the question was, *How do we make sense out of them?* Indeed, another reason for their impact was that they gave new urgency to the longstanding puzzle presented by prosopagnosia: no one denied that the syndrome existed, but a widely accepted explanation had eluded scientists for at least 100 years. This puzzle was articulated by Hans-Lukas Teuber in an anecdote, from the "Perception" section of a chapter co-authored with Brenda Milner (Milner and Teuber, 1968). Stylistic and cultural aspects of this quote will be offensive to contemporary ears, but let's look past the dated language and focus on the substance:

There is the familiar example of perceptual learning, on the part of a white visitor to China (or conversely, a Chinese visiting America). At first, most of the Chinese faces are indiscriminable, but, once one is fairly familiar with a few individual faces, something seems to have been learned that generalizes to all Chinese faces: "Knowing" a few makes all the others more distinguishable. Conceivably, an agnosia for faces could be due to a loss of such an acquired system of rules for classifying facial differences. Alternatively, and even more speculatively, we could

assume that the initial basis for discrimination and identification of faces and of facial expressions involves some innate schemata onto which the subsequent, perceptual differentiations were merely imposed ... This strong but as yet unprovable assumption would make a selective loss of capacity for identifying faces, following certain focal lesions in man, somewhat less paradoxical than it now appears to be. (pp. 300–301)

The concluding sentence of Kanwisher et al. (1997) echoes Teuber's reasoning: "Future studies can also evaluate whether extensive visual experience with any novel class of visual stimuli is sufficient for the development of a local region of cortex specialized for the analysis of that stimulus class, or whether cortical modules like area FF must be innately specified (Fodor, 1983)." (p. 4310)

Perceptual expertise

Many of the "future studies" foreshadowed by Kanwisher et al. (1997) have been led by cognitive neuroscientists Isabel Gauthier and Michael Tarr. (Tarr, like Kanwisher, started his career as a cognitive psychologist and has become increasingly "neuro" as his career evolves.) These two, and their colleagues, have pursued the idea that there's "nothing special" about faces at the neural level – no genetically specified circuits responsible for their processing. Rather, faces are stimuli that we see thousands of times each day, every day, and being able to individuate them is vitally important for success in our highly social environment. Thus, for example, they have argued that it is possible that when Kanwisher et al. (1997) showed different houses vs. different faces in the experiment illustrated in *Figure 9.3* panel 1C, the two types of stimuli may have been processed differently. The house stimuli may have been processed at a more categorical level, the faces at an individual level. (E.g., the subjects may have been thinking "there's another house, and another one, and another one, ...," whereas they may have been scrutinizing each face to determine if it was someone they knew.) By this account, the greater activity for faces than for buildings in that experiment may have had more to do with how stimuli were processed than with what the stimuli were.

This perspective is one that emphasizes the role of expertise in perception. At the neural level, it argues,

FFA may be best construed as a “fusiform expertise area.” (Although no one, to my knowledge, has ever proposed the moniker “FEA,” Tarr and Gauthier (2000) did try rebranding it the “flexible face area.” This never caught on.) Note that what we’ll call the “expertise” account has a ready explanation for why some prosopagnosic patients also lose the ability to recognize other kinds of stimuli with which they have a high degree of expertise, such as the farmer who lost the ability to individuate his cows, or the bird watcher who lost the ability to distinguish among species of bird. Gauthier, Tarr, and colleagues tested this idea with a clever set of studies that entailed training subjects (for some studies healthy young individuals, for others, prosopagnosic patients) to become experts at recognizing individual tokens of a category of stimuli for which there could not exist a genetic template: Greebles.

Greeble expertise

Gauthier and Tarr (1997) developed Greebles as a class of stimuli that have many of the same properties as faces: (1) they are processed configurally – the relation between parts (for faces, distance between eyes, from nose to eyes, etc.) is critically important for recognition; (2) they show an inversion effect – it’s disproportionately difficult, relative to other stimuli, to recognize them when they are upside down; (3) they tend to be classified at the individual level, rather than the category level. They can be categorized at the level of Greeble, family, gender, and individual. As illustrated in *Figure 9.4.*, the fMRI data of Gauthier et al. (1999) suggest that, in contrast to Greeble novices, experts at Greeble recognition evince a statistically comparable stimulus-evoked activity for Greebles as for faces in the FFA. From this, they concluded that “The strongest interpretation suggested by our results ... is that the face-selective area in the middle fusiform gyrus may be most appropriately described as a general substrate for subordinate-level discrimination that can be fine-tuned by experience with any object category” (p. 572).

Here, we have only considered in any detail two publications that can be viewed as early skirmishes in a debate that continues unabated up through the time of this writing. One place to go for updates on this debate is the Web page of Isabel Gauthier’s laboratory, which reads almost like a blog detailing the most recent developments (*Web Link 9.3*).

Evidence for a high degree of specificity for many categories in ventral occipitotemporal cortex

Quickly on the heels of the 1997 publication from Kanwisher and colleagues came a flurry of publications suggesting a further modularization of inferior temporal cortex, some of these from this same group. A *parahippocampal place area* (PPA) was discovered that responds preferentially to viewing a scene (be it a room or a landscape); an *extrastriate body area* (EBA) that responds preferentially to viewing body parts; and a *visual word-form area*. A summary of the location of these areas is provided in *Figure 9.5*. These findings, too, have not been without their skeptics: some, à la Tarr and Gauthier, have pointed to potential confounds in the studies producing the results

FIGURE 9.4 FFA activity in Greeble novices vs. Greeble experts. Source: Gauthier, Tarr, Anderson, Skudlarski, and Gore, 1999. Reproduced with permission of Nature Publishing Group.

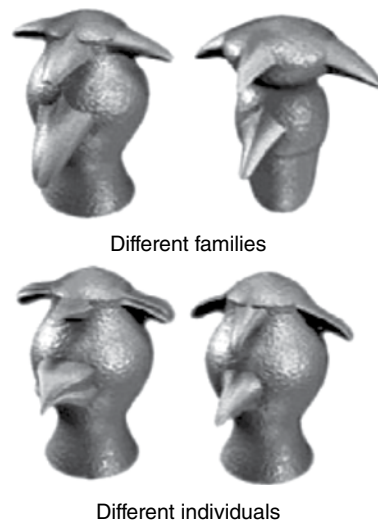


FIGURE 9.4.A Examples of Greebles. The two in the top row are from different “families,” as defined by the large central part. The two at the bottom are two individuals from the same family, differing only in terms of the shapes of smaller parts. Subjects were determined to have achieved expert status with Greebles as quickly as they could two from the same family as quickly as they could two from different families.

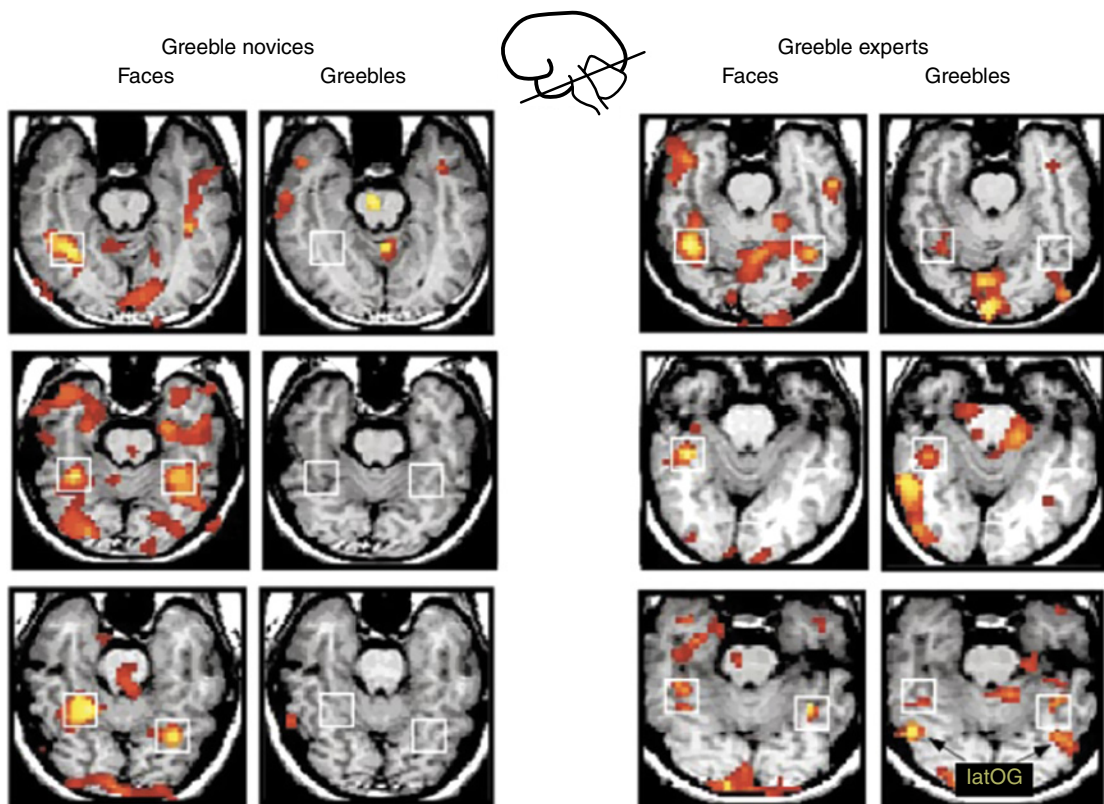


FIGURE 9.4.B Ventral temporal lobe activity, measured with fMRI, from six individuals, during passive viewing of faces vs. Greebles. White squares identify the FFA.

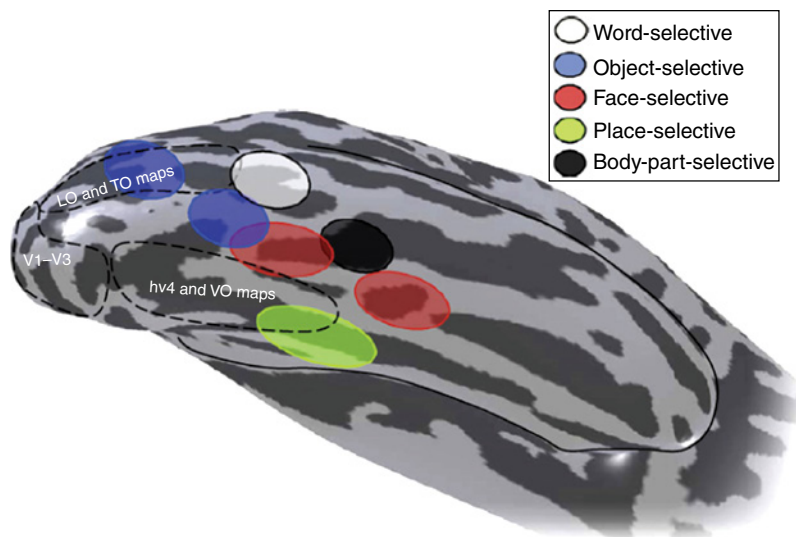


FIGURE 9.5 Juxtaposition of regions on the ventral surface of the occipital and temporal lobes that contain retinotopic maps (dashed lines and black labels) and those that have been identified as preferentially responsive (via univariate analyses) to visual stimuli from various categories. Source: Wandell and Winawer, 2011. Reproduced with permission of Elsevier.

summarized in *Figure 9.5*; others, as we'll see in the next section, have called into question the validity of the assumptions underlying the methods that produce these results.

Evidence for highly distributed category representation in ventral occipitotemporal cortex

Four years after the publication of the Kanwisher, McDermott, and Chun (1997) report on face perception, James Haxby (who we first encountered in *Chapter 6*) and colleagues published a study of object recognition that has also had a remarkable impact on cognitive neuroscience. In contrast to the modular models summarized in the previous paragraph, Haxby et al. (2001) proposed that “The representation of a face or object is reflected by a distinct pattern of response across a wide expanse of cortex in which both large- and small-amplitude responses carry information about object appearance,” and, therefore, that “each category elicits a distinct pattern of response in ventral temporal cortex that is also evident in the cortex that responds maximally to other categories” (p. 2425). To test these ideas, they acquired fMRI while subjects viewed two blocks each of serially presented stimuli drawn from one of each of eight categories: faces, houses, cats, bottles, small man-made tools, shoes, chairs, and smeary “nonsense images.” Next, they compared the similarity of the pattern of activity for each category in its even-numbered block vs. in its odd-numbered block, by correlating the two – that is, they assessed the extent to which knowing the level of signal intensity in each voxel of ventral occipitotemporal cortex during one block predicted what that same voxel's signal intensity was in the other block. (Importantly, they did this regardless of whether any given voxel's value was large or small, positive or negative.) For comparison, they also correlated the pattern produced by the category in question against all of the other categories. (Thus, for each category, eight correlations were performed (one within-category and seven between-category), and for the experiment, eight within-category correlations and 56 between-category correlations.) This is illustrated in *Figure 9.6*, in which, for each row, the red bar represents the within-category correlation coefficient, and the

dark blue the between-category coefficients. (Note that, because the correlations were computed by comparing the values of every voxel in ventral occipitotemporal cortex from one block vs. another, this was a multivariate analysis – indeed, it is considered the grandpappy of the MVPA methods introduced in *Chapter 8*.) The results indicated that coefficients from within-category correlations were higher than those from between-category correlations for 96% of the comparisons. The fact that this analysis was successful for each stimulus category provided support for the authors' proposal that the neural representation of visual categories is broadly distributed.

MVPA challenges models of modular category specificity

Two follow-up analyses went further, and directly challenged the interpretation of studies supporting modular models of the neural organization of categories of information. First, Haxby et al. (2001) redid the analyses described above, but after excluding from the data set the voxels that responded maximally to the critical category. Thus, for faces, for example, they removed the voxels from the FFA before performing the within- and between-category correlations for face blocks. The outcome, as evidenced by comparison of the yellow vs. red bars in *Figure 9.6*, is that whereas the within-category correlation coefficients declined to varying degrees in the “excluded” data sets, they nonetheless remained higher than the between-category coefficients for 94% of comparisons. The implication of this finding is that, for faces, for example, face information is represented in cortex outside of the FFA. This, therefore, argues against the view that faces are processed in a highly modular manner.

But what about the FFA? Even if this region seems not to have “exclusive rights” to the processing of visually presented faces, it could still be “specialized” for face processing. But is it? This question harkens back to the anecdote of Saul Sternberg's question to Nancy Kanwisher about what to make of the fact that FFA does respond, albeit with a smaller magnitude, to stimuli other than faces (*Research Spotlight 9.2*). To address it, Haxby et al. (2001) identified the FFA, the analogous voxels that responded maximally to houses, and so on. Then, using only FFA voxels, for example, they performed within- vs. between-category

correlations for each of the other seven categories. That is, e.g., within the FFA, does the pattern of activity evoked by a block of houses better predict the pattern evoked by the second block of houses than the pattern evoked by bottles, shoes, and so on? The answer was “yes,” for 83% of comparisons. The same was true for six of the analogous ROIs defined with univariate methods as being selective for other categories: the house ROI supported within-category

correlations (for non-house stimuli) at a rate of 93%; the tools ROI at 94%; and the cats ROI at 85%. Only the bottle ROI did not differentially represent non-bottle stimuli. Recall that *Methodology Box 8.2* emphasized the increased sensitivity offered by MVPA over univariate analysis methods. Here are very concrete examples of this fact.

The implication of these findings is that the FFA cannot be characterized as being “specialized,” or

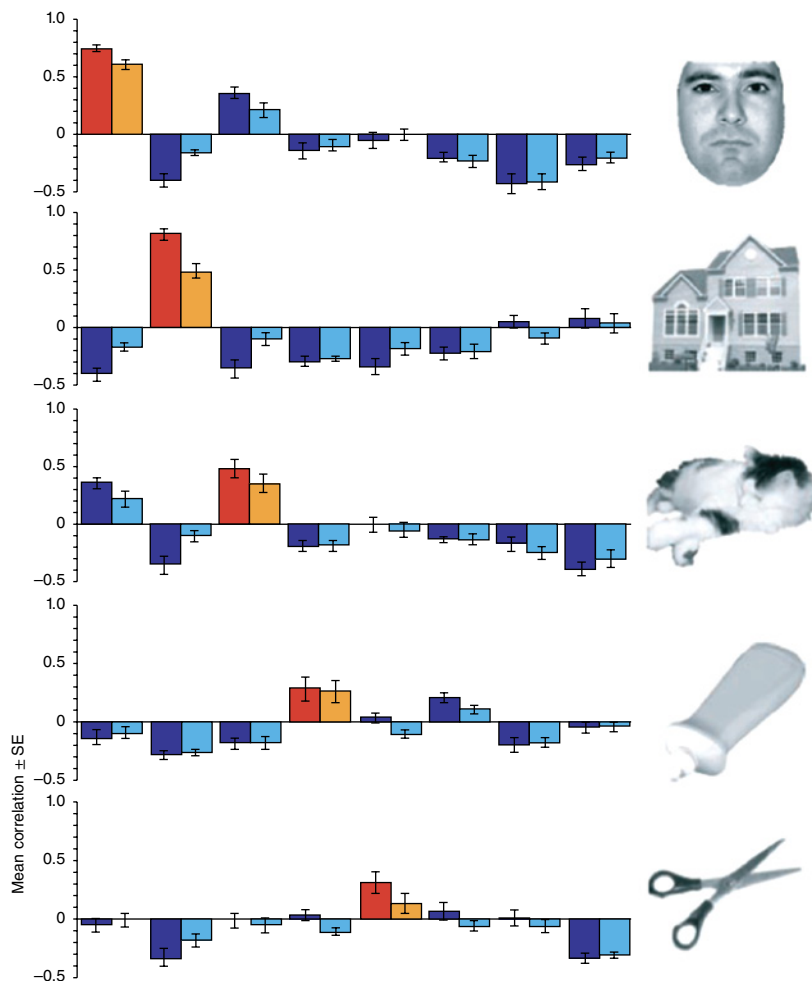


FIGURE 9.6 Correlation matrix showing within- and between-category correlations for the entirety of ventral occipitotemporal cortex that was analyzed, as well as for the same region, but with each category’s maximally responsive voxels removed from the analysis. (The category *tools* is denoted by the scissors.) Source: Haxby, Gobbini, Furey, Ishai, Schouten, and Pietrini, 2001. Reproduced with permission of AAAS.

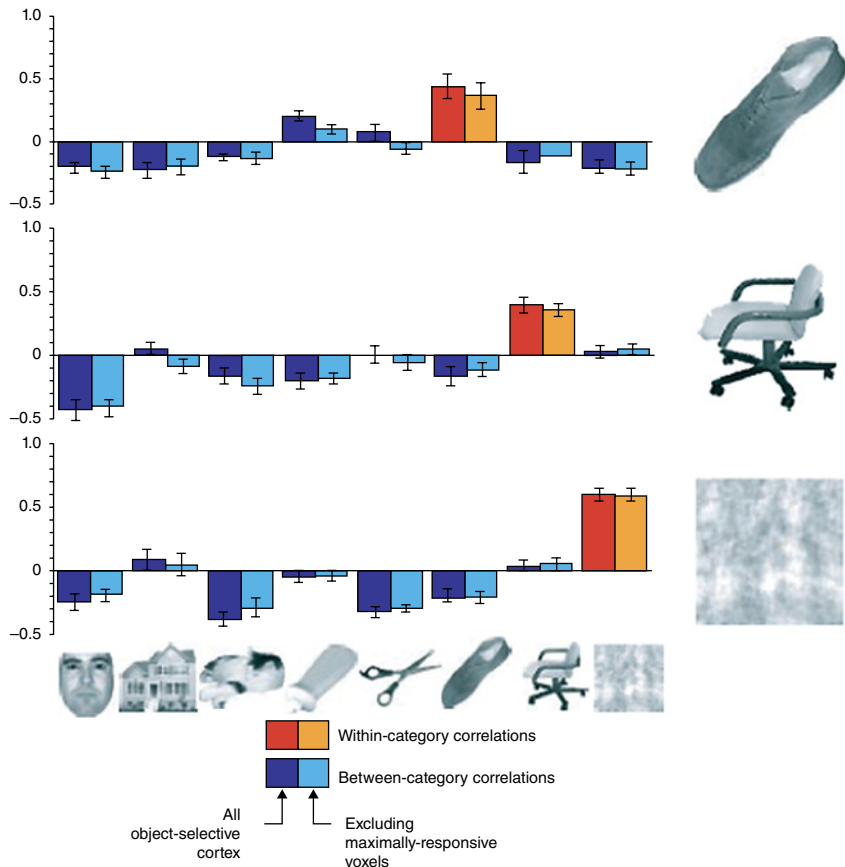


FIGURE 9.6 (Continued)

“selective,” or “specific” for face processing. Is it, however, necessary?

Convincing as the Haxby et al. (2001) analyses were at establishing the distributed nature of object representation in ventral occipitotemporal cortex, it is important to keep in mind that the data were generated in an intact ventral occipitotemporal cortex. With regard to faces, for example, even though one of their analyses found evidence for face representation in cortex outside the FFA (i.e., when the FFA was excluded from the analysis), the tissue of the FFA was intact and functioning normally in the brains of their subjects when the scans were performed. Thus, the Haxby et al. (2001) findings cannot tell us whether non-FFA regions would be able to support a distributed representation of faces if the FFA were not influencing their activity when faces are being presented.

Demonstrating necessity

The caveat at the end of the previous paragraph is a replay of a principle that we have now encountered several times: one cannot know the necessity of a region’s contribution to a function without perturbing it. The final study that we will consider in this chapter employed electrocorticography (ECoG), as well as electrical stimulation via the ECoG grid (*Methodology Box 9.3*), to effect just such a focused perturbation.

The study illustrated in *Figure 9.7* was led by (Azeri-born, US-trained and -based) neurologist Josef Parvizi, in collaboration with (Israeli-born and -educated, US-based) cognitive neuroscientist Kalanit Grill-Spector (a former postdoctoral trainee of Nancy Kanwisher). Although the published report

METHODOLOGY BOX

9.3 Electrocorticography (ECoG)

ECoG refers to the technique of recording electrical signals from the surface of the cortex subcranially and, typically, subdurally. In humans, the procedure is possible for cognitive neuroscience research because the electrodes – typically embedded in a strip that is laid across the surface of the cortex, and sometimes coaxed into sulcal crevasses – have been placed by epilepsy surgeons seeking the source of seizure activity in patients. Thus, experiments are performed while the patient is in the hospital, awaiting surgery to relieve otherwise intractable epilepsy. It is important to note that whereas penetrating, depth electrodes are sometimes also implanted, ECoG signals are more akin to LFP recordings, and thus cannot isolate single-unit, or even multi-unit, activity.

Because Parvizi et al. (2012; *Figure 9.7*) were interested in comparing findings from fMRI with those from ECoG, they focused on oscillatory power in the gamma band, which is believed to be monotonically related to the BOLD signal. (Although this has been demonstrated empirically in simultaneous fMRI and LFP measurements in the monkey (e.g., Logothetis, Pauls, Augath, Trinath, and Oeltermann, 2001), it can also be deduced from what we know about these two signals: the BOLD signal increases with increased metabolic demands driven by synaptic activity, and higher-frequency fluctuations in the LFP result from higher rates of synaptic activity.) The image in panel B of *Figure 9.7* shows the location of two face-specific patches, produced with essentially the same methods as Kanwisher, McDermott, and Chun (1997). (Note from *Figure 9.3* that more than one face-selective patch was often identified in the right hemisphere of ventral occipitotemporal cortex, and this is also codified in the schematic in *Figure 9.5*.) The fMRI scan that produced the image in panel B was acquired prior to the surgical implantation into this same individual of the ECoG array illustrated in panel A, and, although the investigators clearly hoped for such an outcome, it was “mere luck” that, for this patient, two electrodes from the ECoG array fell directly over these two face patches (*Tip 9.10*).

The pie charts in *Figure 9.7 A* represent the results of ECoG recording while the subject viewed stimuli from four categories. Note that these results could, in principle, be interpreted as being consistent with both modularist and distributed accounts of face processing. With regard to the former, electrodes over the fusiform gyrus clearly measure stronger gamma-band responses for faces than for non-face stimuli; with regard to the latter, there are clearly prominent responses to faces, as well as to non-faces, across a large expanse of posterior occipitotemporal cortex from which recordings were obtained. The additional step taken in this study that differentiates it from studies “limited” to ECoG, is that the investigators stimulated the cortex by selectively passing current between pairs of electrodes in the array. The stimulation protocol (2–4 mA delivered in alternating square-wave pulses at 50 Hz) was intended to disrupt normal functioning in the tissue between the two electrodes. In this way, with the same experimental apparatus, the investigators could both localize activity that was correlated with a function of interest, and assess the necessity of the regions so identified by disrupting their activity. (More on recording and analysis of ECoG signals at *Web Clip 9.2*; *Web Clip 9.3* documents the results of the microstimulation experiment.)

contains considerable detail and illustration relating to the ECoG results, the truly novel and consequential finding from this study, produced by the focused electrical stimulation described in *Methodology*

Box 9.3, was so clear and striking that no statistics, indeed, no quantification, were needed (*Tip 9.11*). It was sufficient to post a video clip of the patient’s real-time report of what he was experiencing while

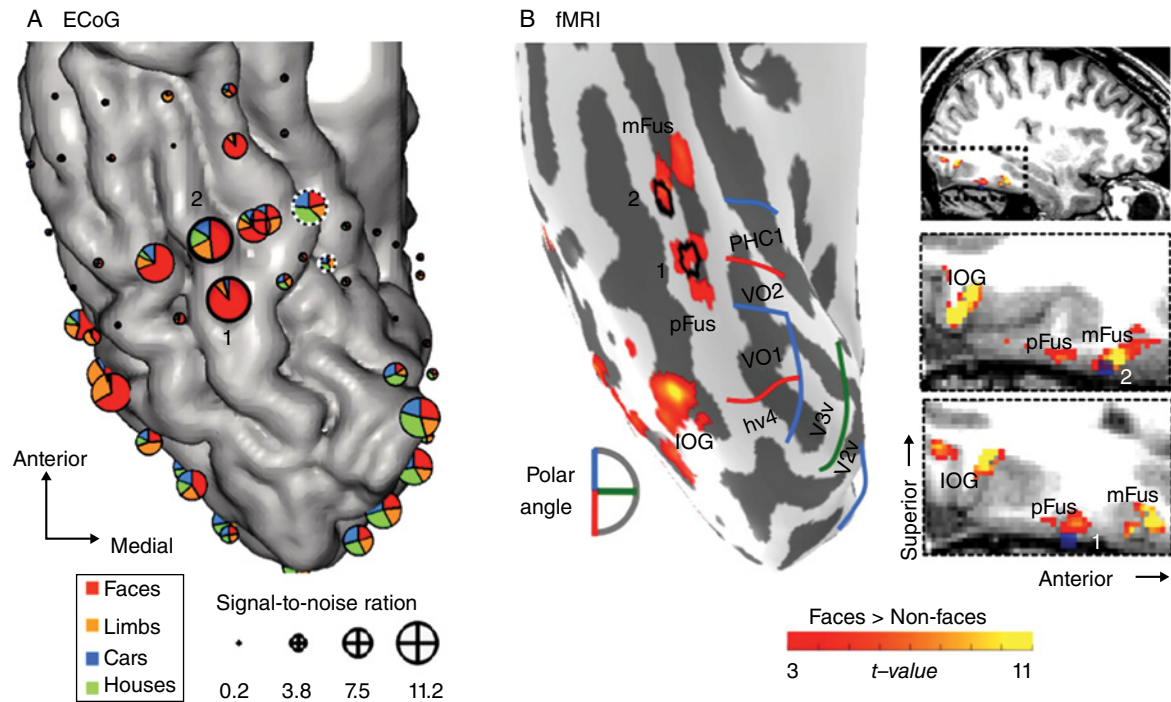


FIGURE 9.7 ECoG and electrical stimulation in the human brain. Panel **A** illustrates the layout of electrodes (black dots and pie charts) on the ventral surface of the patient's posterior right hemisphere. From their placement, one can surmise that two strips of electrodes, each containing two rows, were inserted lateral-to-medial along the ventral surface of the temporal lobe, the more posteriorly positioned one wrapping along the medial wall, and additional strips were positioned more superiorly, lateral and medial to the occipital pole. Each pie chart corresponds to an electrode that responded to one or more of the four categories of visual stimuli, and is color coded to indicate relative selectivity within these four. The size of each pie chart corresponds to the "signal-to-noise ratio," which can be understood as a normalized measure of the stimulus-specific increase in broadband power in the low- and high-gamma range (40–160 Hz) relative to the magnitude of fluctuations at this electrode at baseline. Panel **B** shows different views of regions whose univariate fMRI responses to faces exceeded responses to the other three categories, with colored lines on the inflated surface indicating retinotopically defined boundaries of the regions identified in black type. Superimposed on the inflated image in panel **B** is the location of the electrodes labeled "1" and "2" in panel **A**. Source: Parvizi, Jacques, Foster, Withoft, Rangarajan, Weiner, and Grill-Spector, 2012. Reproduced with permission of the Society for Neuroscience.

Tip 9.10

Another example that, as we already saw with the discovery of orientation-selectivity by Hubel and Wiesel (*Research Spotlight 3.1*), "luck favors the prepared mind."

Tip 9.11

In effect, the authors communicated their key finding via a contemporary implementation, carried out with contemporary technology, of the nineteenth-century practice of performing live demonstrations of the effects of a brain manipulation at a scientific congress (*Chapter 1*)! See for yourself at *Web Clip 9.3*.

looking at Parvizi while stimulation was delivered across electrodes 1 and 2: “‘You just turned into somebody else. Your face metamorphosed.’ When probed further, he reported that features appeared distorted: ‘You almost look like somebody I’ve seen before, but somebody different. That was a trip. ... It’s almost like the shape of your face, your features drooped’” (Parvizi et al., 2012, p. 14918). The patient experienced no such perceptual metamorphoses when he attended to “Get Well Soon” helium balloons in the room, nor when he attended to the television. The video, viewable at *Web Clip 9.2*, is astounding, and leaves no uncertainty about the answer to our question of whether the FFA plays a causal role in face processing.

RECONCILING THE IRRECONCILABLE

Is anyone else dizzy? We’ve covered many profound questions, and in so doing have reviewed many cutting-edge techniques for modeling, measuring, and manipulating the activity of the brain. One take-away from all of this is that seemingly incontrovertible evidence can be mounted for (previously) seemingly incompatible perspectives on how a function like face perception is carried out. Thus, it is objectively the case that the FFA is not specialized for face representation, but it is also clearly the case that it makes necessary contributions to the normal performance of this function. Stay tuned ...

END-OF-CHAPTER QUESTIONS

1. What factors call into question the idea that associative agnosias may reflect truly unimpaired perception “stripped of its meaning”?
2. How can prosopagnosic patients successfully navigate our modern, intensely social, society?
3. Do PDP models of visual object recognition “prove” that there is no distinction between “structural” and “semantic” levels of visual representation?
4. What was it about the Kanwisher, McDermott, and Chun (1997) results that some viewed as contrary to how they thought that the brain works?
5. What are the alternative explanations of the Kanwisher et al. (1997) results proffered by Gauthier, Tarr, and colleagues?
6. In what sense do the findings of Haxby et al. (2001) rule out a strong version of the argument that the FFA is specialized for face processing?
7. In what sense do the findings of Parvizi et al. (2012) establish that the FFA makes necessary contributions to face processing?

REFERENCES

- Benson, D. Frank, and John P. Greenberg. 1969. “Visual Form Agnosia: A Specific Defect in Visual Discrimination.” *Archives of Neurology* 20 (1): 82–89. doi: 10.1001/archneur.1969.00480070092010.
- Eggert, Gertrude H., ed. 1977. *Wernicke’s Works on Aphasia: A Sourcebook and Review*. The Hague: Mouton.
- Farah, Martha J. 1990. *Visual Agnosia*. Cambridge, MA: MIT Press.
- Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fuster, Joaquín M. 2003. “More Than Working Memory Rides on Long-Term Memory.” *Behavioral and Brain Sciences* 26 (6): 737–737. doi: 10.1017/S0140525X03300160.
- Gauthier, Isabel, and Michael J. Tarr. 1997. “Becoming a ‘Greeble’ Expert: Exploring Mechanisms for Face Recognition.” *Vision Research* 37 (12): 1673–1682. doi: 10.1016/S0042-6989(96)00286-6.
- Gauthier, Isabel, Michael J. Tarr, Adam W. Anderson, Pawel Skudlarski, and John C. Gore. 1999. “Activation of the Middle Fusiform ‘Face Area’ Increases with Expertise in Recognizing Novel Objects.” *Nature Neuroscience* 2 (6): 568–573. doi: 10.1038/9224.
- Haxby, James V., M. Ida Gobbini, Maura L. Furey, Alumi Ishai, Jennifer L. Schouten, and Pietro Pietrini. 2001. “Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex.” *Science* 293 (5539): 2425–2430. doi: 10.1126/science.1063736.

- Humphreys, Glyn W., and M. Jane Riddoch. 1987. "The Fractionation of Visual Agnosia." In *Visual Object Processing: A Cognitive Neuropsychological Approach*, edited by Glyn W. Humphreys and M. Jane Riddoch, 281–307. Hove, UK: Lawrence Erlbaum Associates.
- Kanwisher, Nancy. 1991. "Repetition Blindness and Illusory Conjunctions: Errors in Binding Visual Types with Visual Tokens." *Journal of Experimental Psychology: Human Perception and Performance* 17 (2): 404–421. doi: 10.1037/0096-1523.17.2.404.
- Kanwisher, Nancy, Josh McDermott, and Marvin M. Chun. 1997. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception." *Journal of Neuroscience* 17 (11): 4302–4311. <http://www.jneurosci.org/content/17/11/4302.full.pdf+html>.
- Logothetis, Nikos K., Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. 2001. "Neurophysiological Investigation of the Basis of the fMRI Signal." *Nature* 412 (6843): 150–157. doi: 10.1038/35084005.
- Milner, Brenda, and Hans-Lukas Teuber. 1968. "Alteration of Perception and Memory in Man: Reflections on Methods." In *Analysis of Behavioral Change*, edited by Lawrence Weiskrantz, 268–375. New York: Harper and Row.
- Parvizi, Josef, Corentin Jacques, Brett L. Foster, Nathan Withoft, Vinitha Rangarajan, Kevin S. Weiner, and Kalanit Grill-Spector. 2012. "Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception." *Journal of Neuroscience* 32 (43): 14915–14920. doi: 10.1523/JNEUROSCI.2609-12.2012.
- Plaut, David C. 2002. "Graded Modality-Specific Specialisation in Semantics: A Computational Account of Optic Aphasia." *Cognitive Neuropsychology* 19 (7): 603–639. doi: 10.1080/02643290244000112.
- Pohl, Walter. 1973. "Dissociation of Spatial Discrimination Deficits Following Frontal and Parietal Lesions in Monkeys." *Journal of Comparative and Physiological Psychology* 82 (2): 227–239. doi: 10.1037/h0033922.
- Rogers, Timothy T., John R. Hodges, Matthew A. Lambon Ralph, and Karalyn Patterson. 2003. "Object Recognition Under Semantic Impairment: The Effects of Conceptual Regularities on Perceptual Decisions." *Language and Cognitive Processes* 18 (5–6): 625–662. doi: 10.1080/016909603440000053.
- Rogers, Timothy T., Matthew A. Lambon Ralph, Peter Garrard, Sasha Bozeat, James L. McClelland, John R. Hodges, and Karalyn Patterson. 2004. "Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation." *Psychological Review* 111 (1): 205–235. doi: 10.1037/0033-295X.111.1.205.
- Tarr, Michael J., and Isabel Gauthier. 2000. "FFA: A Flexible Fusiform Area for Subordinate-Level Visual Processing Automatized by Expertise." *Nature Neuroscience* 8 (3): 764–769. doi: 10.1038/77666.

OTHER SOURCES USED

- Bukach, Cindy M., and Jessie J. Peissig. 2010. "How Faces Become Special." In *Perceptual Expertise: Bridging Brain and Behavior*, edited by Isabel Gauthier, Michael J. Tarr, and Daniel Bub, 11–40. Oxford: Oxford University Press.
- Ellis, Hadyn D. 1996. "Bodamer on Prosopagnosia." In *Classic Cases in Neuropsychology*, edited by Chris Code, Claus-W. Wallesch, Yves Joanette, and André Roch Lecours, 69–75. Hove, UK: Psychology Press.
- Foster, Brett L., and Josef Parvizi. 2012. "Resting Oscillations and Cross-Frequency Coupling in the Human Posteromedial Cortex." *NeuroImage* 60 (1): 384–391. doi: 10.1016/j.neuroimage.2011.12.019.
- Gross, Charles G. 1994. "Hans-Lukas Teuber: A Tribute." *Cerebral Cortex* 4 (5): 451–454. doi: 10.1093/cercor/4.5.451.
- Pribram, Karl H. 1977. "Hans-Lukas Teuber: 1916–1977." *American Journal of Psychology* 90 (4): 705–707. <http://www.jstor.org/stable/1421744>.
- Riddoch, M. Jane, and Glyn W. Humphreys. 1987. "A Case of Integrative Visual Agnosia." *Brain* 110, 1431–1462.
- Rubens, Alan B., and Frank Benson. 1971. "Associative Visual Agnosia." *Archives of Neurology* 24 (4): 305–316.

FURTHER READING

- Çukur, Tolga, Shinji Nishimoto, Alexander G. Huth, and Jack L. Gallant. 2013. "Attention During Natural Vision Warps Semantic Representation across the Human Brain." *Nature Neuroscience* 16 (6): 763–770. doi: 10.1038/nn.3381. A "mind-warping" demonstration of a state-of-the-art (as of this writing) fMRI investigation of the neural bases of visual knowledge representation.

Farah, Martha J. 1990. *Visual Agnosia*. Cambridge, MA: MIT Press.

An encyclopedic synthesis of 100 years of research on the topic; be sure to look at the most recent edition, as it will contain updates on developments in this fast-moving literature since the original 1990 publication.

Fodor, Jerry A. 1985. "Précis of *The Modularity of Mind*." *Behavioral and Brain Sciences* 8 (1): 1–5. doi: 10.1017/S0140525X0001921X.

A quicker read than the 1983 monograph, plus, it is supplemented by interesting critiques from others in the field, followed by Fodor's rejoinder.

Logothetis, Nikos K. 2003. "The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal." *Journal of Neuroscience* 23 (10): 3963–3971.

A review summarizing the relation between the LFP and BOLD signals that trades accessibility for some of the technical detail packed into Logothetis et al. (2001).

Wandell, Brian A., and Jonathan Winawer. 2011. "Imaging Retinotopic Maps in the Human Brain." *Vision Research* 51(7): 718–737. doi: 10.1016/j.visres.2010.08.004.

If you didn't look this one up after reading Chapter 2, here's another prompt! This paper presents a thorough summary of neuroimaging studies of the functional organization of the ventral stream, including category representation.

CHAPTER 10

NEURAL BASES OF MEMORY

KEY THEMES

- Foundational knowledge about the organization of memory, and the amnesic syndrome, has resulted from the pioneering research of Brenda Milner and her associates with the patient H. M.
- The model assembled from the global amnesic syndrome is one of dissociable systems supporting declarative vs. nondeclarative vs. short-term memory, and of dissociable processes of encoding, consolidation, storage, and retrieval.
- Long-term memory (LTM) is underlain by the principle of association, and an influential model of a neurobiological and computational mechanism that might implement associative learning in the brain was proposed by D. O. Hebb in 1949.
- Hebbian learning seems to be embodied in the phenomenon of long-term potentiation (LTP) of synaptic strength, discovered by Lømo and Bliss in the 1960s and 1970s.
- A critical implementer of many types of LTP is the NMDA subtype of glutamate receptor, which acts as a synaptic coincidence detector.
- As of today, there are many proposals for “how the hippocampus works,” but not yet any broad consensus on which one is likely to be right.
- The phenomena of memory consolidation, and the more-recently characterized reconsolidation, pose important challenges for cognitive neuroscience, but also potential therapeutic opportunities.

CONTENTS

KEY THEMES

PLASTICITY, LEARNING, AND MEMORY

THE CASE OF H. M.

Bilateral medial temporal lobectomy

The global amnesic syndrome

The organization of memory

The subjective experience of amnesia

Hippocampus vs. MTL?

ASSOCIATION THROUGH SYNAPTIC MODIFICATION

The example of Pavlovian conditioning

Hebbian plasticity

Long-term potentiation

*Hippocampal circuitry and its
electrophysiological investigation*

*The NMDA glutamate receptor as coincidence
detector*

Multiple stages of synaptic strengthening

The necessity of NMDA channels for long-term
memory formation

HOW MIGHT THE HIPPOCAMPUS WORK?

Fast-encoding hippocampus vs. slow-encoding
cortex

Episodic memory for sequences

Episodic memory as an evolutionary elaboration
of navigational processing

WHAT ARE THE COGNITIVE FUNCTIONS OF THE HIPPOCAMPUS?

Standard anatomical model

Challenges to the standard anatomical model

Newer lesioning methods

Perceptually challenging behavioral tasks

"New roles" for the hippocampus?

Consolidation

TO CONSOLIDATE

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

PLASTICITY, LEARNING, AND MEMORY

In several previous chapters, most notably when we considered use-dependent somatosensory reorganization, motor learning, and the simulation of learning by adjusting connection weights in a parallel distributed processing (PDP) model, we have touched on a fundamental property of the nervous system, which is its ability to modify itself in order to refine its representation of the outside world. Here, finally, we will address the phenomenon of *neural plasticity*, and its behavioral and theoretical counterparts *memory* and *learning*, head-on. We'll begin with a summary of an influential neurosurgical case that revealed many principles of human learning and memory, then consider computational and physiological models of cellular and molecular bases of these phenomena, and conclude at the systems level. This will provide a foundation for the subsequent three chapters' explorations of the kinds of memory that we use every day: episodic, semantic, and working.

THE CASE OF H. M.

Whether in terms of raw citation counts or subjective assessment of breadth and depth of legacy, H. M. may well be the single most influential case in the history of cognitive neuroscience. He spent his life in the vicinity of Hartford, CT, where, in 1953, at the age of 27, he underwent a neurosurgical procedure that truly shaped the course of scientific history.

Bilateral medial temporal lobectomy

In the back seat of his parents' car, on the day of his fifteenth birthday, Henry Molaison was stricken by his first **grand mal epileptic seizure** (*Tip 10.1*). Over the next few years his neurological condition progressively worsened, to the point where he dropped out of high school, couldn't hold down a steady job, and, by his mid 20s, was effectively housebound. For several years he and his family had consulted with a neurosurgeon at Hartford Hospital,

Tip 10.1

Throughout his life, Henry was identified in scientific presentations and publications by his initials. Upon his death in 2008, with no living immediate relatives, his anonymity was unmasked.

Dr. William Scoville, who, after exhausting other treatment strategies, proposed an experimental surgical intervention. Because EEG had failed to identify a clear focus of Henry's seizures, Scoville proposed the surgical removal of the hippocampus, known to be the most epileptogenic tissue in the brain. (A subsequent section summarizing the circuitry of the hippocampus will provide some insight into why this may be.)

A diagram of the procedure, the surgeon's estimate of the extent of removal, and MR images from the 1990s, are illustrated in *Figure 10.1*. Although the surgery was successful from the narrow perspective that it brought H. M.'s epilepsy under control, its unexpected consequences were profound and permanent: H. M. lost the ability to consciously remember anything that he experienced from that time forward, for the remaining 50+ years of his life.

The global amnesic syndrome

The influence of H. M.'s case is due in large part to the pioneering research of Dr. Brenda Milner, a British-born and -trained neuropsychologist who has spent the majority of her scientific career at McGill University and the Montreal Neurological Institute, working with such luminaries as Wilder Penfield (who we met in *Chapter 4*) and Donald Hebb (who we will meet further along in this chapter). Shortly after H. M.'s surgery, Scoville came across a report from Penfield and Milner detailing rare instances of profound memory impairment resulting from unilateral removal of the medial temporal lobes, and speculating that this may have resulted from the fact that the unoperated hemisphere was also, unbeknownst to them, dysfunctional. Scoville contacted Penfield to tell him about his case and, soon thereafter, Milner traveled from Montreal to Hartford to initiate her historic series of studies. This research

established several facts about the consequences of bilateral damage of the medial temporal lobe (MTL):

- Within the realm of cognition, its effects are by-and-large restricted to memory. Thus, for example, H. M.'s ability to understand and produce language was unaffected by his surgery, as was his IQ, which actually increased marginally after his surgery, no doubt due to the reduction of epileptic symptoms.
- Within the realm of memory, its effects are primarily restricted to the acquisition of new memories, in that facts acquired and events occurring before the damage remain accessible. Thus, bilateral damage of the MTL produces an *anterograde amnesia* (loss of memory going forward in time).
- The comparatively less severe *retrograde amnesia* is limited to the few years immediately prior to the insult – remote memories are spared.
- Within the realm of memory acquisition, the deficit is limited to events and facts that can be consciously recalled; patients can continue to learn new skills and habits that don't require conscious recollection.
- Also spared by MTL damage is short-term memory, the ability to remember a telephone number while looking for your phone or, as considered in *Chapter 8*, to briefly remember the location previously occupied by a stimulus that is no longer there.

Some of these facts have been refined and/or called into question in the ensuing years, but at this early juncture in the chapter it will be useful to summarize them as principles whose tenets, across this and the next three chapters, we can then test and revise as appropriate. (*Web Clip 10.1* offers “Snapshots of H. M.’s contributions to the science of memory.”)

FIGURE 10.1 H. M.’s lesions.

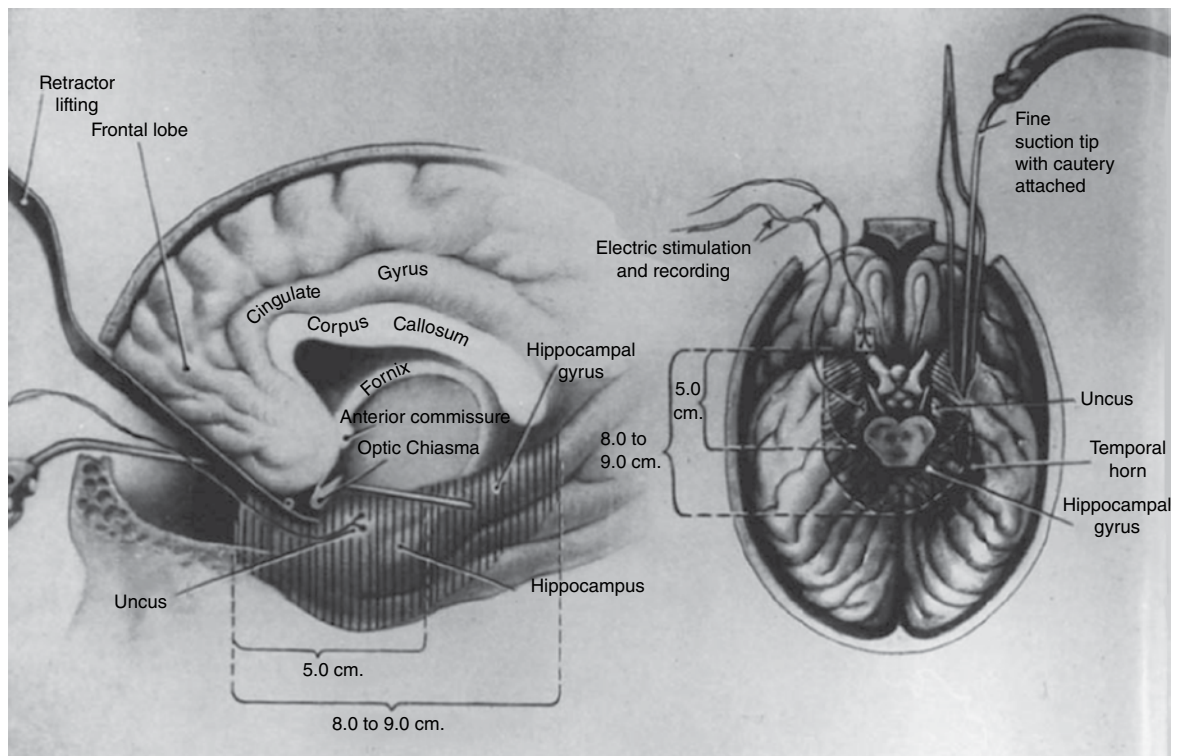


FIGURE 10.1.A Illustration of the surgical procedure, and estimated extent of removal, of the hippocampus. Source: Scoville and Milner, 1957. Reproduced with permission of BMJ.

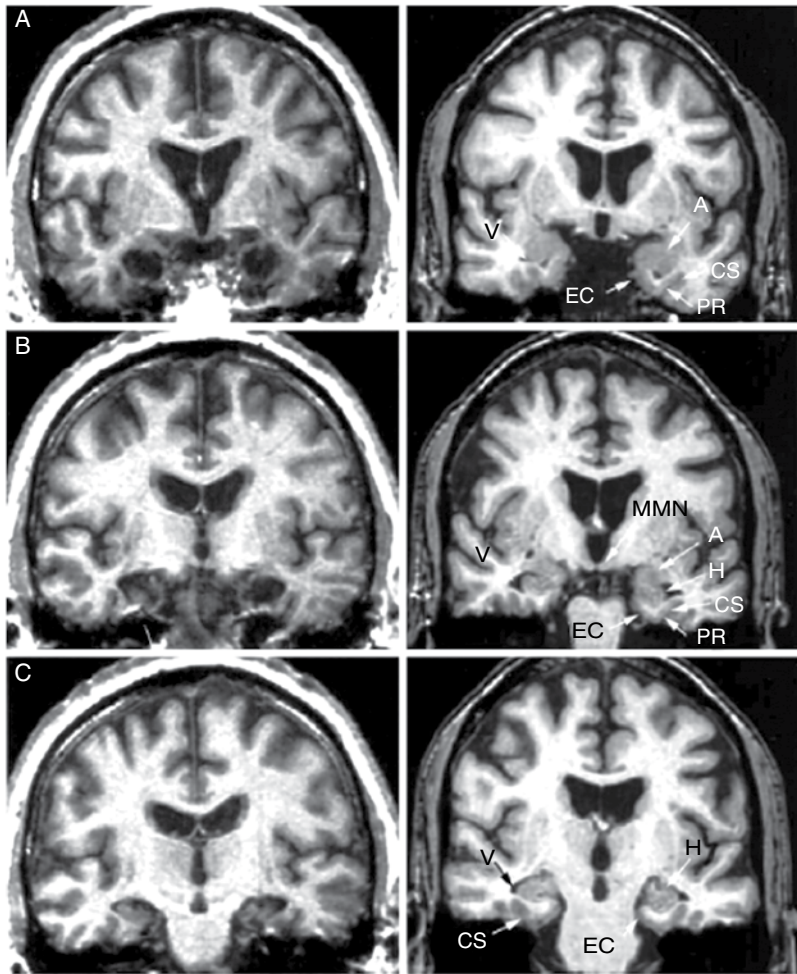


FIGURE 10.1.B T1-weighted MR images of the brain of H. M. (left-hand column) and an age- and education-matched control subject, at three progressively more caudal locations. Labels on the control images identify landmarks and regions that are damaged or missing in H. M. (A=amygdala; CS=collateral sulcus; EC=entorhinal cortex; PR=perirhinal cortex; MMN=medial mammillary nuclei; H=hippocampus; V=ventricle). At the caudal-most slice, atrophied remnants of hippocampal tissue that weren't removed by the surgery are visible. MMN, not directly affected by the surgery, are reduced in size in H. M., presumably due to the loss of afferent drive that they would typically receive from the hippocampus, via the [fornix](#) (see panel **A**). Source: Adapted from Corkin, Amaral, González, Johnson, and Hyman, 1997. Reproduced with permission of the Society of Neuroscience.

The organization of memory

The principles arising from the work of Milner and colleagues include:

- (1) There is a fundamental distinction between [declarative memory](#) (or [explicit memory](#)) and [nondeclarative memory](#) (or [implicit memory](#)). The

former refers to consciously accessible records of facts and events and depends on the integrity of the MTL. The latter includes any example of experience-dependent plasticity that modifies a behavior in a manner that does not require conscious access for its expression, and includes conditioning, motor learning, and perceptual learning.

(2) Declarative and nondeclarative memory are both considered long-term memory (LTM), and are distinct from short-term (or working) memory (STM).

(3) There is a neuroanatomical basis for dissociating the processes of encoding, storage, and retrieval in declarative memory. In the case of H. M., as presented so far, his deficit seems to be one of **encoding** – processing newly encountered information such that it can later be remembered. If either storage or retrieval were compromised, he wouldn't be able to retrieve memories from prior to his surgery. Thus, storage and retrieval would not seem to depend on the MTL.

(4) Long-term declarative memories undergo a period of **consolidation**, to which the MTL contributes importantly, after which they become relatively immune to disruption by MTL damage.

The subjective experience of amnesia

While the case of H. M. and of many other patients have been invaluable for the scientific study of memory, they also raise profound questions that get at the core of what it means to be a sentient human being, and confront us with heartbreakingly poignant conditions before which, at the present time, we are helpless (*Thought Box 10.1*).

Hippocampus vs. MTL?

At the time of the initial findings with H. M., and for most of the ensuing quarter century, most of the findings and principles summarized here were framed in terms of the hippocampus (e.g., “declarative memory is hippocampally dependent”). Subsequently, there has been an upswing in interest in the contributions of the neocortical regions of the MTL – the entorhinal, perirhinal, and parahippocampal cortex – to memory functions, and the extent to which each may make distinct contributions. First, though, let's give some consideration to what it means to create and store a memory.

ASSOCIATION THROUGH SYNAPTIC MODIFICATION

The example of Pavlovian conditioning

What is a memory? Fundamentally, it is the association of two previously unlinked pieces of information. This concept was introduced in *Research Spotlight 9.1*, but

here we'll be more concrete, with a cartoon example familiar to most. In the early twentieth century, the Russian physiologist Ivan Pavlov (1849–1936), whose research on the physiology of digestion garnered the Nobel Prize in 1904, noticed that dogs in his laboratory would often begin salivating in advance of the arrival of their food. Formal experimentation established that the animals could be taught to associate an arbitrarily selected stimulus (let's say the ringing of a bell) with the act of getting fed. The essence of the protocol was to ring the bell at the time of food delivery. After several such trials, it would be sufficient to ring the bell, alone, and the dog would begin to salivate. In a strict sense the dog is expressing a memory, because prior to the experiment the ringing of a bell did not produce this behavior. Therefore, it is the influence of its recent experience that is being expressed in its behavior. If one wants to project human-like cognition onto the dog, one can speculate that hearing the bell makes the dog think of being fed (i.e., makes it *remember* being fed).

Staying at this schematic, cartoon level, what must have happened within the brain of our subject? Prior to the experiment, the sound of a bell would activate its auditory system, and would perhaps trigger an exogenous shift of attention, but certainly no activity in the digestive system. During training, this auditory circuit was active at the same time as were the olfactory circuits that processed the smell of the food and then activated what we'll call the “digestive neurons” that trigger the salivary reflex. Something about the coactivation of these two circuits led to a strengthening of the link between the auditory representation of the bell and the digestive neurons. In effect, the dog's brain perceived an association in the world that hadn't existed previously, and modified its own connectivity to represent it. Thus, at this neural level, the memory “is” this new connection. Now, let's drill down to consider how this might come to pass.

Hebbian plasticity

If one were to rank universities in terms of their influence on memory research, one could make a strong case for placing McGill University in Montreal at the top. It was from here that, in the same year that Brenda Milner began her graduate studies under his supervision, the Canadian Psychologist Donald Hebb (1904–1985)

THOUGHT BOX

10.1 What is it like to be amnesic?

Self report from H. M. suggests that one phenomenological quality of anterograde amnesia is a pervasive anxiety about what may have happened just beyond the edge of the present moment:

Right now, I'm wondering, have I done or said anything amiss? You see, at this moment everything looks clear to me, but what happened just before? ... It's like waking from a dream; I just don't remember ... Every day is alone, in itself. Whatever enjoyment I've had, whatever sorrow I've had. (Hilts, 1996, p. 138).

On another occasion, during an exchange between H. M. and researcher William Marslen-Wilson (relayed by Hilts, 1996), the patient confessed to worrying about giving the wrong answer, whether during formal testing or just in conversation:

It is a constant effort, Henry said. You must always "wonder how is it going to affect others? Is that the way to do it? Is it the right way?" ... Asked if he worried about these things a lot, struggled with his thought to get right answers, he said yes, all the time. "But why?" "I don't know," said Henry. (p. 140)

The anxiety and frustration of amnesia is even more evident in Clive Wearing, a distinguished British musicologist, whose amnesia resulted from herpes encephalitis. This condition can produce severe damage to the hippocampus while leaving the rest of the brain relatively unscathed (*Tip 10.2*). In a clip from a television documentary (viewable at [Web Link 10.1](#)), he and his wife are sitting in a city park:

WIFE:	Do you know how we got here?
WEARING:	No.
WIFE:	You don't remember sitting down?
WEARING:	No.
WIFE:	I reckon we've been here about 10 minutes at least.
WEARING:	Well, I've no knowledge of it. My eyes only started working now ...
WIFE:	And do you feel absolutely normal?
WEARING:	Not absolutely normal, no. I'm completely confused.
WIFE:	Confused?
WEARING (AGITATEDLY):	Yes. If you've never eaten anything, never tasted anything, never touched anything, never smelled something, what right have you to assume you're alive?
WIFE:	Hmm. But you are.
WEARING:	Apparently, yes. But I'd like to know what the hell's been going on!

(In making these pronouncements about his senses, it is clear that Wearing is speaking figuratively, not literally.) Thus, for Wearing, too, each waking moment feels as though he is just waking up from sleep. The journal that he keeps is filled with multiple entries that all contain variants of the same message. For example, directly under the entry "10:49am I AM TOTALLY AWAKE – FIRST TIME," which appears on the first line of a page, is a second entry "11:05am I AM PERFECTLY AWAKE – FIRST TIME," and so on. When left alone in his room, the patient fills entire pages in this way with entries made at intervals ranging from 5 to 45 minutes.

These anecdotes capture an essential quality of the conscious phenomenology of the MTL amnesic patient, the near-continual impression of just having awakened from unconscious sleep. The plight of the MTL amnesic patient, then, is to be fully cognizant of, if not preoccupied by, the fact that one is not cognizant of the daily events of one's life.

Tip 10.2

Among the regions damaged by H. M.'s surgery were his amygdala, a structure of central importance to the processing of emotion (*Chapter 17*). As a result, another unintended byproduct of his surgery was that his mood was often blunted.

published his book *The Organization of Behavior* (1949), which articulated what has come to be the single most influential idea in the history of memory research: “When an Axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased” (p. 62). This has come to be known as “Hebb’s postulate.” For two decades it existed as “just” a compelling idea, until its empirical confirmation in the 1960s and 70s. Since that time, it has been fundamental for understanding synaptic plasticity in biological systems, and is the basis for synaptic modification in virtually all computational neural network models, including PDP models that were introduced in the previous chapter.

Before we consider the physiological correlates of Hebb’s postulate, let’s apply it to our Pavlovian conditioning example. What it is saying is that there was something special about the fact that bell-related activity in the auditory system coincided with the activity in the digestive neurons. Thus, even though the auditory activity initially had no causal role in driving the digestive neurons, something about the co-incidence of their activity led to a strengthening of the ability of auditory activity to drive salivary activity. To know what that something was, we’ll next turn to experiments initiated in Oslo in the mid-1960s.

Long-term potentiation

Hippocampal circuitry and its electrophysiological investigation

For our fanciful trip to Oslo to be most profitable, we’ll need to be familiar with the circuitry of the hippocampus. As illustrated in *Figure 5.5*, the hippocampus can be construed as sitting atop the cortical hierarchy, receiving highly processed inputs

Tip 10.3

Whereas to some anatomists the shape of the hippocampus suggested a seahorse lying on its side (see *Figure 2.2*), to others it suggested the ram’s horn of the Egyptian god Ammon. The result for the nomenclature is that the former names the structure, the latter (*cornu ammonis*) gives its initials to two of the structure’s prominent fields.

Tip 10.4

From this view, one might think that the hippocampus, like the cerebellum, is organized as a bank of parallel, encapsulated circuits. Unlike the cerebellum, however, the third dimension is critically important to the function of the hippocampus, in that collaterals from CA3 run the long axis of the structure, thereby implementing widely divergent processing within the hippocampus itself.

from the highest levels of the dorsal stream (via feedforward projections from PPC to the parahippocampal cortex (PHC)) and the ventral stream (via feedforward projections from IT and STP to perirhinal cortex (PRC)). PHC and PRC each funnel into entorhinal cortex (EC), which then projects to hippocampus. The hippocampus is considered the top of the hierarchy because its projections don’t go to yet another downstream area. Instead, after the incoming information is processed through its circuitry, the pattern of cortical convergence onto the hippocampus is inverted, and hippocampal outputs are fed back out to EC, then divergently to many regions of neocortex.

The hippocampus itself differs from six-layer neocortex in that it is, in effect, a rolled-up sheet of three-layer archicortex (“old” cortex). *Figure 10.2.A* shows the hippocampus in cross section, *10.2.B* how, within this 2-D slice, one can trace a complete circuit from the inputs from EC to the dentate gyrus (DG), through the CA fields (*Tip 10.3*), then back out to cortex (*Tip 10.4*). This circuitry lends itself to various experimental

FIGURE 10.2 The hippocampus in cross section.

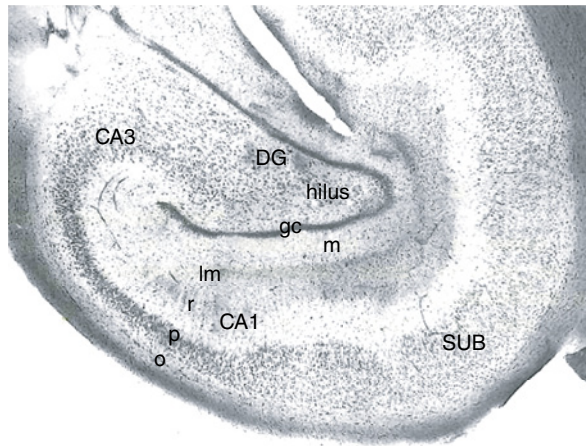


FIGURE 10.2.A Coronal section through right hemisphere. DG = dentate gyrus, and within the DG, gc = granule cell layer, and m = stratum moleculare. The layers of the CA fields are: lm = stratum lacunosum-moleculare; r = stratum radiatum; p = stratum pyramidale; o = stratum oriens. SUB = subiculum, an output station not included in *Figure 10.2.B*. Source: Buszáki, 2006. Reproduced with permission of Oxford University Press, USA.

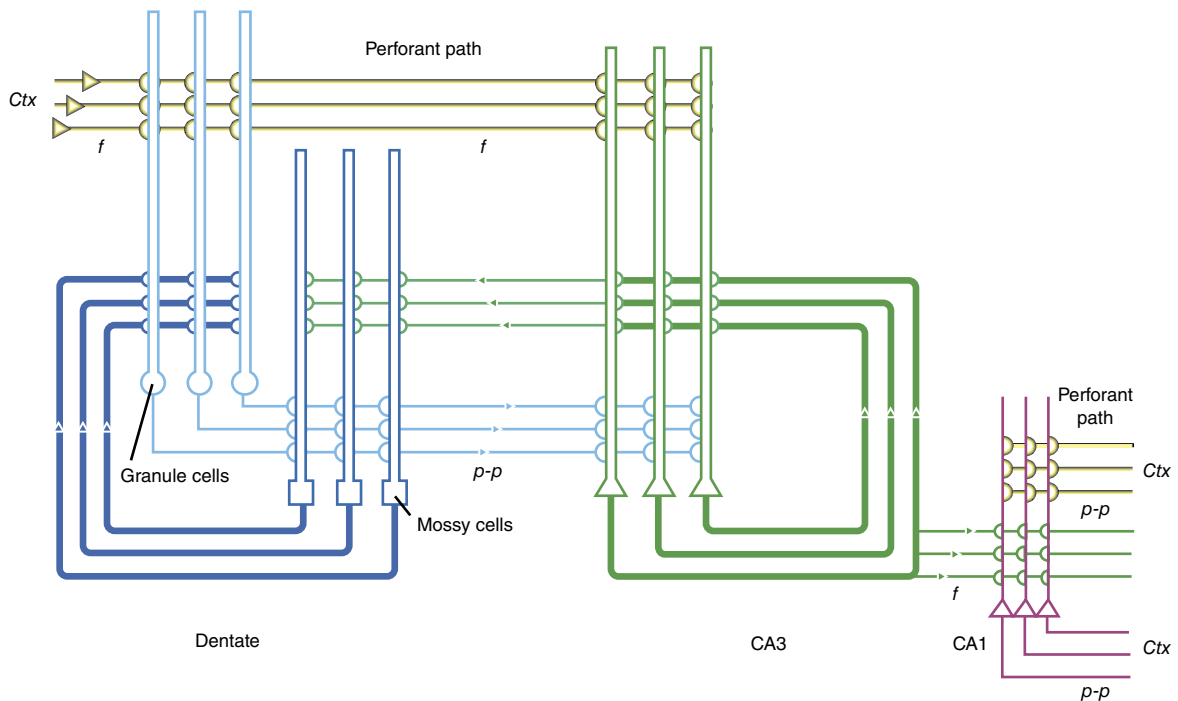


FIGURE 10.2.B A simplified wiring diagram of the circuitry in the slice if *Figure 10.2.A* were unrolled. The perforant path projects to two regions of hippocampus, the DG, and CA1. The projections from EC to DG are “fanning” (i.e. divergent), as are those from CA3 to CA3 (i.e., “recurrent” projections) and to CA1, whereas those from EC to CA1 and from DG to CA3 are “point-to-point” (p-p). Source: Lisman, 1999. Reproduced with permission of Elsevier.

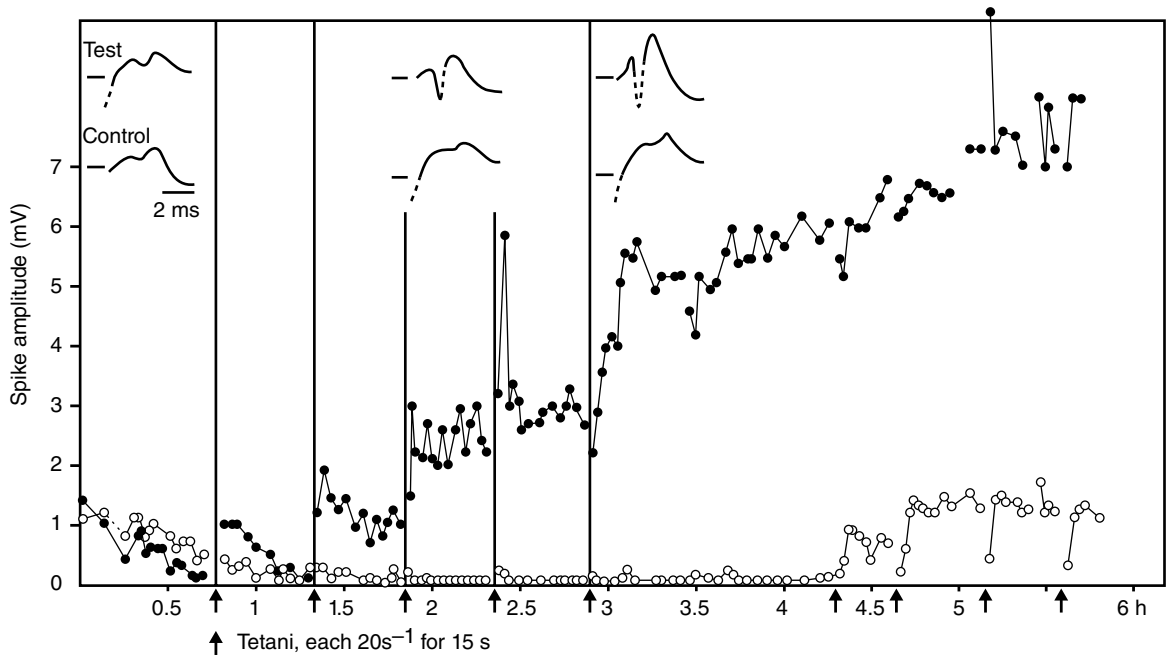


FIGURE 10.3 Long-term potentiation (LTP) in the dentate gyrus of the hippocampus (for details, see *Research Spotlight 10.1*). Source: Lømo, 2003. Reproduced with permission of the Royal Society.

procedures that wouldn't be feasible in other brain regions. Now we can drop in on the University of Oslo in the mid-1960s, where the electrophysiologist Per Andersen had pioneered the measurement of the “population spike” in the DG. Due to the homogeneity and relative feedforward simplicity of the circuitry, if one delivers electrical stimulation to the axons projecting from EC to the DG (the “performant path”), a field potential recorded from the granule cell body layer of the DG gives a straightforward measure of the aggregate strength of the connection between neurons in these two regions (*Figure 10.3*).

The breakthrough finding relating to Hebbian plasticity was first made by Andersen's (Norwegian) graduate student Terje Lømo, and later confirmed and codified by Lømo in collaboration with British neurophysiologist Timothy Bliss. The essence of their findings is summarized in *Research Spotlight 10.1*.

There is near-universal consensus that LTP, as illustrated in *Figure 10.3*, is a core phenomenon underlying synaptic plasticity in many regions of the brain, and that it is fundamental to learning and memory (*Tip 10.5*). However, although we've now seen how LTP can be

Tip 10.5

The converse type of plasticity, long-term depression (LTD), also plays a critical role in learning and memory. It is the basis, for example, of the plastic changes in the cerebellum that we reviewed in *Chapter 7*.

induced, we still haven't addressed how it happens. How does a synapse strengthen itself? To do this, we have to move to the *in vitro* slice preparation.

The NMDA glutamate receptor as coincidence detector

Another convenient feature of hippocampal circuitry for neurophysiologists is that it is relatively straightforward to remove a cross-sectional “slice” of the hippocampus from a live animal (most typically a rodent) and keep it alive *in vitro* by perfusing it with tissue culture medium. Such a slice will contain intact circuitry as illustrated in *Figure 10.2*, and its electrophysiology and pharmacology can be studied with many

RESEARCH SPOTLIGHT

10.1 The discovery of long-term potentiation (LTP)

The protocol of Bliss and Lømo (1973) was to deliver high-frequency (20 Hz) trains (15 seconds in duration) of electrical pulses (“tetani”) to the perforant path of an anaesthetized rabbit, and measure the resultant population spike from the granule cell-body layer of the DG. As illustrated in *Figure 10.3*, prior to delivery of the first tetanus, they measured the baseline strength of the EC-DG connection by delivering single pulses. The waveforms at the top of the figure show the baseline response, a negative-going – here, negative is plotted up – potential with two bumps, that then returns to baseline. Why is negative plotted up? The negative-going potential corresponds to the depolarization of granule cells (intracellular depolarization causing extracellular hyperpolarization, which is what the extracellular electrode records), the “dimple” in the middle of the depolarization wave is the population spike (after discharge, a neuron rapidly hyperpolarizes, producing a positive shift in the extracellular potential). At baseline, very few DG neurons generate an action potential in response to single probing pulses to the perforant path, and so the population spike is very small. Each data point along the horizontal axis indicates the magnitude of the population spike produced by a single probing pulse (responses to the tetani themselves are not shown), and the baseline population spikes from time 0 to 0.75 h (i.e., 45 min.) are between 1 and 0 mV.

At time 0.75 h, a tetanus is delivered to the perforant path in the “test” hemisphere, and the population spike to subsequent probing pulses in this hemisphere is briefly elevated (filled circles), whereas the control hemisphere is largely unchanged. At time 1.35 h, after the second tetanus, the response of the population spike is larger, and only weakens slightly, leveling off at 1 mV over the ensuing 30 min. At time 1.85 h, after the third tetanus, the population spike is further strengthened to 2.5 mV. (Representative waveforms from this epoch are illustrated at the top of the figure.)

Finally, just before time 3 h, a final tetanus is delivered. From this point forward, the population spike starts to grow “on its own,” without any further tetani, from a magnitude of ~3 mV to in excess of 7 mV. (The four tetani delivered after time 4 h were delivered to the control hemisphere.) Critically, the potentiated state of the population spike was sustained for the next three hours, until the experiment was terminated. Ergo, “long-term potentiation.” (After six hours of recording, preceded by several hours of set-up and surgery, it was surely time to go to the pub!)

From these results it was inferred (and later confirmed) that this strengthening of the population spike reflected Hebbian plasticity, because it came about after the initial pulses of the tetanus had depolarized the granule cells, such that later pulses in the tetanus-generated action potentials. (Delivery of the same number of single pulses, separated in time, would not have this effect.) Thus, it was the co-incidence of (upstream) perforant path activity and (downstream) granule cell firing that resulted in the strengthening (the “potentiation”) of the synapse.

techniques that are either not possible, or much more difficult, in an in vivo preparation. Indeed, much of what we know about the cellular and molecular bases of LTP has come from study of the synapse of CA3 axons on CA1 dendrites, a synapse that is much easier to study in the slice than in vivo. This research has revealed that there are many variants of LTP, and many factors that

influence it, only a few of which we can review here. Our focus will be on a prominent form of Hebbian LTP, one that is observed at the EC-DG synapse and also at the CA3-CA1 synapse: NMDA-dependent LTP.

As illustrated in *Figure 10.4*, there are two major classes of glutamate receptor, each named after a compound that is a selective agonist. (AMPA receptors

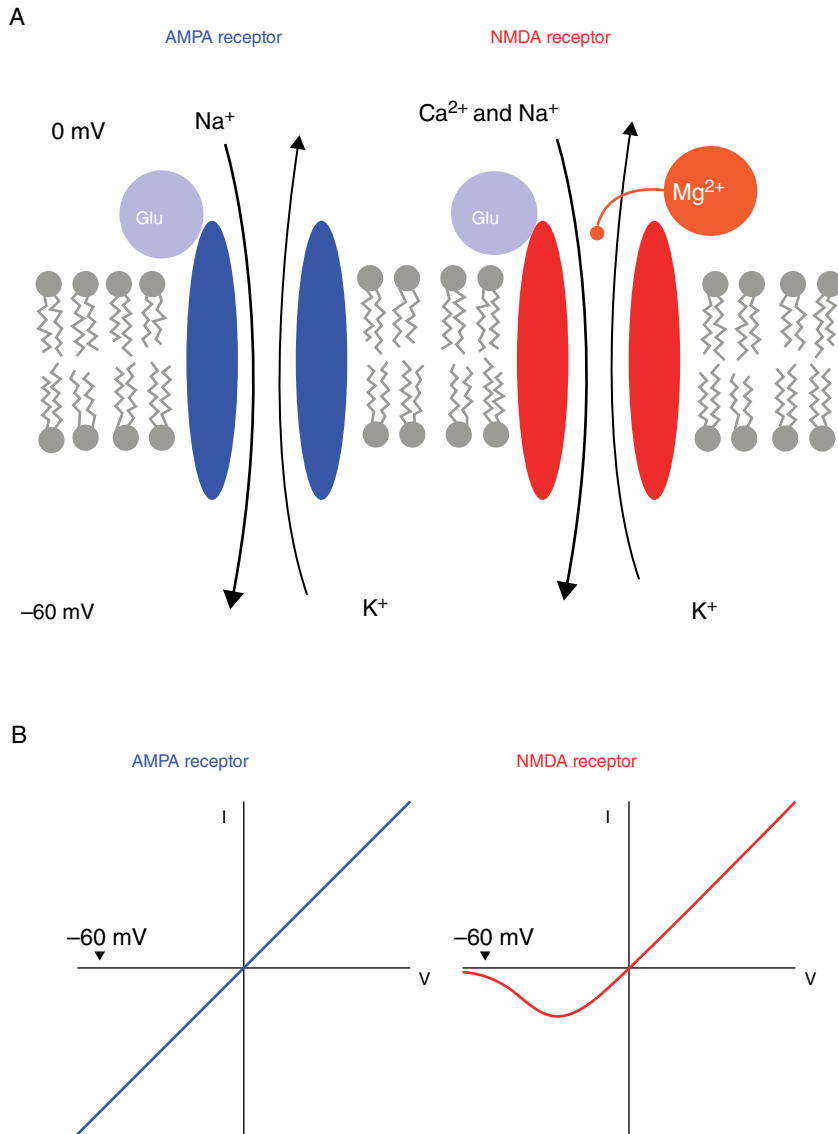


FIGURE 10.4 The two major classes of glutamate receptor. **A.** The NMDA receptor features a Mg^{2+} ion in its pore, which is dislodged when (1) glutamate is bound to its binding site, and (2) a state of depolarization inside the synaptic terminal repels the positively charged Mg^{2+} ion. **B.** The current–voltage (I – V) relationship of the two, illustrating that the NMDA receptor only passes current when the neuron is depolarized. Note that these plots follow the convention of labeling a flow of negative charges as positive current, a consequence of the fact that electrons are negatively charged. Therefore, the negative value of I when the membrane voltage is negative corresponds to an influx of positively charged ions. At the cell's resting potential, the current passed by bound NMDA receptors is effectively 0. When the cell's potential depolarizes, however, to between -60 and 0 mV, the bound NMDA receptor passes a negative current (i.e., an influx of positively charged ions). Source: Lüscher and Malenka, 2012. Reproduced with permission of Cold Spring Harbor Laboratory Press.

received considerable treatment in *Chapter 2*; if you're hazy on the details, now would be a good time to review them.) The NMDA receptor differs from the AMPA receptor in that its pore is occupied by a magnesium ion (Mg^{2+}). As a result, binding by glutamate, which untwists its pore, is not sufficient for the NMDA receptor to allow the influx of positively charged ions, because the Mg^{2+} ion remains in place, blocking the pore. To activate the NMDA receptor, there must be a co-incidence of two events: (1) presynaptically released glutamate binds to its binding site; and (2) the postsynaptic terminal is already depolarized. A depolarized state of the postsynaptic terminal has the effect of repelling the positively charged Mg^{2+} ion, thereby forcing it out of the pore, which allows for the influx of Na^+ and, more importantly, Ca^{2+} . As discussed previously, Ca^{2+} acts as a catalyst for many intracellular chemical cascades, and in the postsynaptic terminal, those cascades lead to LTP.

Before detailing the events triggered by the influx of Ca^{2+} , however, let's take another moment to emphasize the functional role of the NMDA receptor. It acts as a coincidence detector, in that it can only be activated by the co-incidence of two events that implement Hebb's postulate: (1) release of glutamate by the presynaptic terminal of the synapse that is to be strengthened; and (2) a prior state of depolarization of the postsynaptic terminal. To make this concrete, we can return once again to our Pavlovian conditioning vignette. The presentation of food activates olfactory neurons, which, in turn, activate digestive neurons. This provides the postsynaptic depolarization. If, during this time, the bell is rung, glutamate released by auditory neurons onto their synapses with digestive neurons will activate NMDA receptors, due to the co-incidence of olfactory activity and auditory activity. This activation of NMDA receptors, as we're about to see, can trigger LTP.

Multiple stages of synaptic strengthening

Finally, we can now explicitly consider the events that lead to the long-term potentiation of synapses. The influx of calcium ions through NMDA receptors has both short-term and long-term consequences. In the short-term, it triggers the insertion of more AMPA receptors into the postsynaptic density. (As it turns out, postsynaptic terminals are equipped for rapid strengthening, with "racks" of AMPA receptors at the ready, to be inserted into the postsynaptic density as needed.) Because the binding of neurotransmitter with a receptor

is a stochastic event, adding more AMPA receptors to the postsynaptic density will necessarily increase the probability that any one molecule of glutamate will bind with a postsynaptic receptor. The aggregate effect is that each action potential has a stronger influence on the postsynaptic neuron – the synapse has been potentiated.

In the long term, Ca^{2+} -initiated cascades of intracellular signaling pathways initiate the synthesis of new proteins that are necessary for new synapses to be constructed. Electron-microscopic images have documented the appearance of new dendritic spines, presumed to be forming new synapses with efferent axons, a so-called structural form of LTP.

The necessity of NMDA channels for long-term memory formation

Since the pioneering work of Lomo and Bliss, the study of LTP has exploded into a subfield of its own. *What molecules are necessary for or influence its expression? Are there presynaptic, as well as postsynaptic changes underlying it? Are there non-Hebbian, non-NMDA-dependent variants? What is the importance of the precise timing of depolarization of the postsynaptic cell and the firing of the presynaptic cell?* Indeed, the enthusiasm for the investigation of LTP has prompted a skeptical backlash from some scientists who have questioned the importance of LTP in the grand scheme of understanding learning and memory. Some examples from this vibrant, dynamic literature are listed in the *Further Reading* section. Before concluding our focus on Hebbian plasticity as a basis for memory formation, *Research Spotlight 10.2* will summarize an elegant study demonstrating the necessity of the NMDA receptor for mammalian spatial LTM.

HOW MIGHT THE HIPPOCAMPUS WORK?

I had an argument with myself (as H. M. used to say) about whether to use the word "does" or "might" in this section heading. In the end, I've concluded that "does" would imply that this question is settled to a degree that it simply is not. What follows, however, are some of the more intriguing and influential possibilities.

RESEARCH SPOTLIGHT

10.2 The necessity of NMDA receptors for spatial LTM

This study, by British neuroscientist Richard Morris and colleagues, is perhaps the most famous experiment to have used the experimental apparatus of his invention, the “Morris Water Maze.” It is a circular pool (~2 m in diameter) filled with water that has been made opaque by the addition of milk. At one location is a small platform that is located just below the surface of the water. The water is kept chilly, to guarantee that when an already water-averse rat is placed in the pool, it will seek to climb up onto the platform ASAP. Because the platform cannot be seen, however, on the first trial of an experiment the rat must swim around and around until bumping into it. On subsequent trials, it can draw on its memory for where the platform had been on previous trials, thereby escaping faster.

To assess the importance of NMDA receptors for spatial learning, Morris, Anderson, Lynch, and Baudry (1986) administered to rats the NMDA antagonist AP5. Importantly, previous studies in the slice had established that although AP5 blocks LTP formation, it does not influence normal synaptic transmission. Together with two control groups, the rats were given 15 trials, across four days, during which to learn where the platform was located. On the critical test day, the platform was removed from the pool, and the rats made to swim for 60 seconds.

As illustrated in *Figure 10.5*, an intact rat caught in this diabolical quandary will swim back-and-forth over the location where it remembers the platform to have been. The AP5 animals, however, just swam around in circles, showing no bias for the location where the platform had been in its 15 previous trips to the pool. To quantify this,

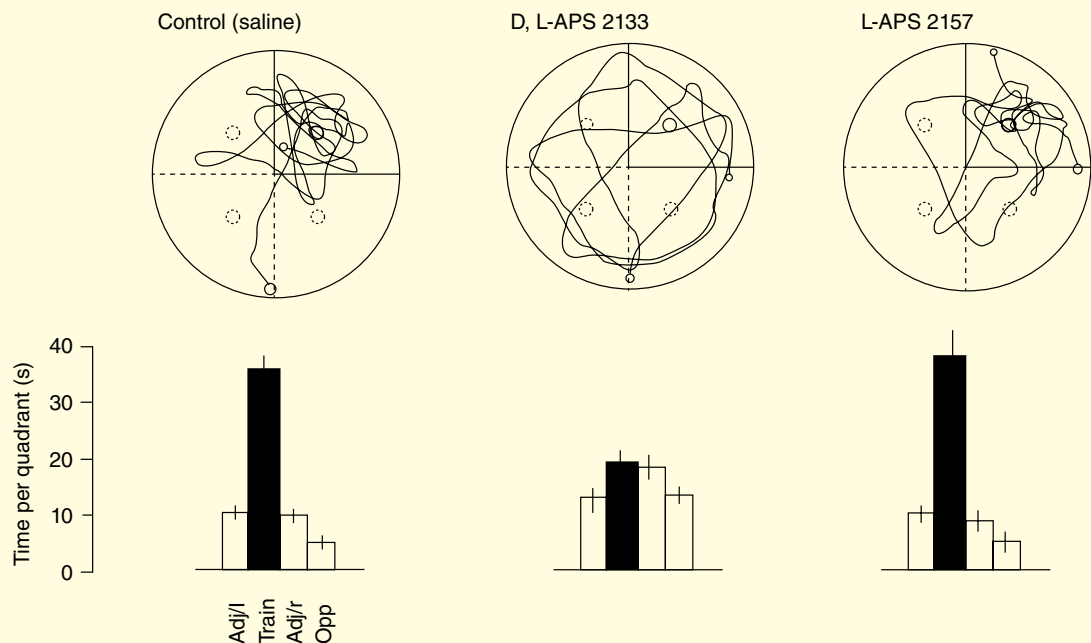


FIGURE 10.5 NMDA receptor blockade blocks spatial LTM in the rat. Top row shows the swimming trajectory on the test day of an animal in each of the three groups. Dotted circles indicate the location in each quadrant analogous to where the platform had been in the critical quadrant (top right for these three illustrations). Bottom row shows the group-average time spent in each of the four quadrants of the pool, labeled according to their position relative to the platform (“Train”). Source: Morris, Anderson, Lynch, and Baudry, 1986. Reproduced with permission of Nature Publishing Group.

Research Spotlight 10.2 *Continued*

the investigators divided the pool into four imaginary quadrants, one centered on the location where the platform had been. The “time per quadrant” plots show that both the saline control animals and the animals treated with a chemically ineffective variant of AP5 (L-AP5) spent a disproportionate period of their 60 seconds in the remembered quadrant, whereas the group treated with the active variant of the compound, “D, L-AP5,” did not.

An electrophysiological study replicating the procedure of Bliss and Lomo (1973) confirmed a total blockade of LTP in the D, L-AP5 group relative to the two control groups.

Finally, a control study also provides insight into a distinctive feature of rodent learning and memory that will be relevant further along in this chapter. It was designed to rule out the possible confound that AP5 may have disrupted sensorimotor and/or motivational processes, rather than spatial LTM. It consisted of two platforms protruding above the water, one homogenous gray, the other black-and-white striped. One was stable and provided escape, the other was floating and sank each time the rat tried to climb on. Black curtains surrounded the pool, and the location of the two platforms was changed on each trial, making this a nonspatial, object memory task (think back to Pohl, 1973). The result was that all groups learned at the same rate. This is consistent with the view that, in the rodent, hippocampal LTP may be specialized for *spatial* learning and memory (*Tip 10.6*).

Fast-encoding hippocampus vs. slow-encoding cortex

The “complementary learning systems” model was formulated by a collaboration between two PDP modelers, James McClelland and his then-student Randall O’Reilly, and hippocampal neurophysiologist Bruce McNaughton. Their model was motivated by two primary considerations, one empirical and one theoretical. The empirical consideration was the evidence for a time-limited role for the hippocampus in LTM storage. That is, the fact that retrograde memory loss is typically limited to memories acquired for the few years prior to the insult suggests either a transfer of memories from the hippocampus to cortex or a time-limited role for the hippocampus in the consolidation of cortically based long-term memories. The theoretical consideration is as follows. There are two primary functions carried out by LTM: the retention of memories for specific episodes that we have experienced; and the extraction from life experiences of general patterns about how the world works. The former need to be highly detailed, and discriminable from other episodes that may have some overlapping features. (E.g., there are two neighbors who frequently walk their dogs by my house, both neighbors are middle-aged white women, both of their pets are Portuguese water dogs. Missou likes to be petted, Ralph

tried to bite me the last time I approached him.) These are called **episodic memories**. The latter need to discard details that are idiosyncratic to any one specific episode and extract general principles. (E.g., dogs are pets that can’t be house-trained, and so need to be walked so that they can relieve themselves outdoors. It is safe to pet some dogs but not others; therefore, it’s best to ask the walker before engaging with a dog.) These are called **semantic memories**, and correspond to knowledge about the world that does not require reference to the episode(s) in which it was acquired.

Computational modeling has demonstrated that the formation of distinct episodic memories, an operation sometimes referred to as “pattern separation,” is computationally incompatible with the abstraction across specific episodes to distill and/or update general principles, an operation sometimes referred to as “pattern completion.” In particular, the rapid representation and retention of arbitrary associations, a hallmark of episodic memory, cannot be supported by a system that learns general principles by abstracting across multiple individual episodes, a necessarily slow process. The complementary learning systems model of McClelland, McNaughton, and O’Reilly (1995), therefore, proposes that the hippocampus serves as the special-purpose rapid learner of episodic memories, and that the way in which these memories are periodically reinstated

(i.e., in full detail, or just the important principle that it contains) determines whether or not it will gradually transfer to cortex as an episodic or a semantic memory. (*Web Clip 10.2.A* provides a more in-depth look at pattern separation, and *Web Clip 10.2.B* considers how the age-related decline of LTM may relate to this process.)

Episodic memory for sequences

Figure 10.2 illustrates that the circuitry of the hippocampus includes two recurrent loops of circuitry: in addition to feeding forward to CA3, granule cells of the DG also project to mossy cells of the DG which, in turn, project back onto the granule cells; and, in addition to feeding forward to CA1, CA3 neurons also feedback onto other CA3 neurons (as well as back onto mossy cells). Indeed, a critical detail that is not captured in this diagram is that individual CA3 axons vary in length from an astounding 150 to 400mm, and each makes between 25,000 and 50,000 synapses with other neurons! This makes of the CA3 layer a “giant random connection space, a requisite for combining arbitrary information” (Buszáki, 2006, p. 285). That is, it is an ideal substrate for the rapid formation of episodic memories invoked by McClelland, McNaughton, and O’Reilly (1995). The neurophysiologist and theoretician John Lisman has emphasized that this recurrent circuitry, combined with the direct EC-to-CA1 projection of the perforant path, enables the hippocampus to represent two kinds of information that are critical for episodic memory: the sequence in which events occurred, and the context in which they took place.

An empirical finding that underlies Lisman’s (1999) idea comes from the phenomenon of phase precession of the firing of hippocampal place cells, discovered by John O’Keefe, who you’ll remember from *Chapter 6* as the original discoverer of place cells. The phenomenon is cartooned in *Figure 10.6(a)*. Lisman’s explanation is that, because the ion channels of NMDA receptors in the CA3 and CA1 regions remain open for 100ms or longer, the recurrent circuitry in this region can encode (via LTP) the temporal association of “location *B* comes after location *A*,” and so on. Importantly, these associations are asymmetrical: although NMDA receptors at the *A*-to-*B* synapse can detect the “lagged coincidence” that *A was recently active*, and *B is now firing*, at the *B*-to-*A* synapse, the

reverse cannot happen, because when *B* is firing, *A* is no longer depolarized. (Thus, although glutamate released from *B* might untwist the pores of NMDA channels in *A*, the absence of depolarization in *A* means that the Mg^{2+} blockade will remain in place.) Lisman further posits that if there is a buffer that can sustain activity in a representation over several seconds, either within the hippocampus or elsewhere, this same mechanism could also encode the sequential order of events spaced out by seconds, rather than tens of milliseconds (we’ll consider data consistent with this idea from human fMRI studies in *Chapter 11*). Note that regardless of the actual time elapsed between two sequentially occurring events, the representation of their ordinal relation would be compressed into a time scale clocked by the theta- and gamma-band oscillations in the hippocampus. This is illustrated in the fictional event cartooned in *Figure 10.6(c)* in which, while at the zoo, (1) *Jerry dropped his candy*, (2) *the monkey in the cage grabbed it*, and (3) *Jerry was sad*. Although these three facts are seemingly activated simultaneously while one thinks about this event, the order in which they occurred would be encoded in the phase of the theta cycle at which the ensemble representing each of these three propositions fires.

Context, according to Lisman (1999) might be carried by perforant path signals from EC directly to CA3. Explicit memory recall, by this scheme, would play out as follows. First, neurons of EC that represent the context of “the zoo” depolarize CA3 neurons at a subthreshold level. Second, the subset of these CA3 neurons that also receive input from the DG representing *Jerry dropped his candy* would be depolarized to their firing threshold, and their activity would represent *Jerry dropped his candy at the zoo*. This activity, would, in turn, activate the CA3 representation of the Puckish monkey who snatched away the candy, and so on. (*Web Clip 10.3* provides more on the role of the hippocampus in memory retrieval.)

Episodic memory as an evolutionary elaboration of navigational processing

For the final model of hippocampal function that we’ll consider here, it will be useful to bring to mind the place cell data from *Chapter 6*, as well as *Figure 6.15’s*

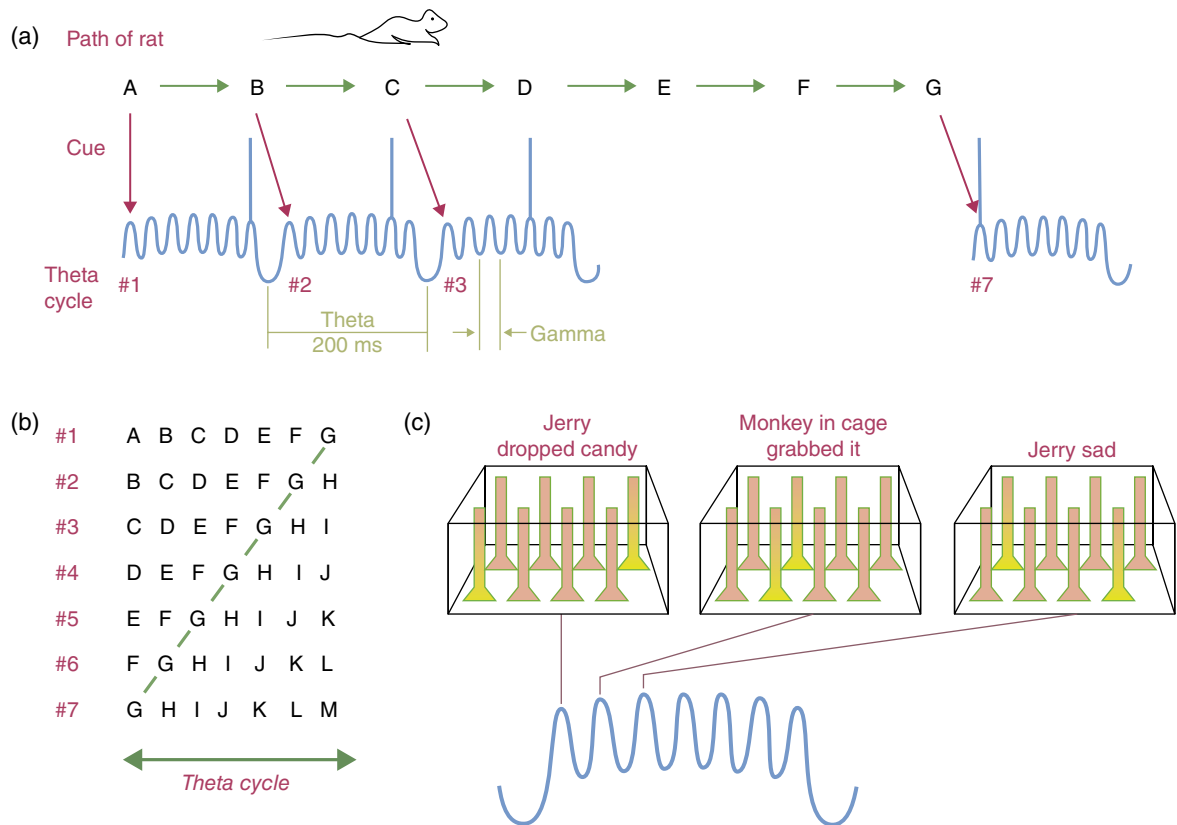


FIGURE 10.6 Empirically documented and theoretically posited representation of sequence in the hippocampus. (a) illustrates the phenomenon of phase precession of place cell firing. As a rat runs from location A to location G, the place cell representing location G initially fires late in the phase of the theta cycle, but its spike precesses to progressively earlier in the phase as the animal approaches G. Each spike occurs at the peak of the gamma-band oscillations that “ride on top of” each theta cycle. (b) illustrates how on each theta cycle, as the rat moves toward G, the “G cell” fires earlier. (c) is the example that Lisman (1999) uses to illustrate how the same principle may underlie LTM for sequences of events. Source: Lisman, 1999. Reproduced with permission of Elsevier.

illustration of the dorsal location of the hippocampus in the rodent. At the level of implementation it draws on many of the same mechanisms as Lisman’s memory-for-sequences model, differing primarily in the emphasis on “how our memory came to work the way that it does.” An influential account by the (Hungarian-born and -trained, US-based) neurophysiologist György Buszáki uses as a starting point the fact that hippocampal place cells encode allocentric representations of space, and thus would be an effective substrate for one-dimensional (1-D) “dead reckoning” navigation (a.k.a., “route following”). 1-D maps, like that illustrated in Figure 10.6(a), would be sufficient to represent an

episodic memory, just as Lisman (1999) illustrates in Figure 10.6(c). If one adds together a sufficient number of 1-D routes one can construct a 2-D map, and the CA3 field as a whole can be construed as a 2-D map, with distances encoded in synaptic weights. (I.e., by the same logic that Lisman uses to argue that these cells can encode asymmetric sequences of events, one would also expect stronger synaptic links between place cells representing nearby locations than between those representing distant locations.) Indeed, Buszáki notes, CA3 probably holds many, many maps that are anatomically intermingled, thus obscuring any topography that may exist for any one map. Such 2-D maps would

be used in “landmark-based” navigation (a.k.a., “map reading”), a more efficient form of navigation that has been observed in many animal species, as well as in humans. Such 2-D maps, Buszáki argues, may be the neural basis of explicit memories. Thus, the hippocampus, a region phylogenetically older than the neocortex, may have originally evolved to support navigation, foraging, and caching behaviors in our vertebrate ancestors. As the neocortex in early mammals acquired the ability to represent information other than just what have been called the “four Fs” – *feeding, fighting, fleeing, and ... mating* – it may have coopted the substrate already in place in the hippocampus for the encoding and storage of information other than space.

Compatible with this “navigational origins” view is the “hippocampus as cognitive map” perspective championed by the cognitive neuroscientist Lynn Nadel, co-author with John O’Keefe of the influential 1978 book of the same name. In a recent review, Nadel, Hoscheidt, and Ryan (2013) add an additional level of detail by drawing on the fact that the size of hippocampal place fields in the rodent changes along a gradient from small in the dorsal to relatively large in the ventral hippocampus. They infer from this that in the rodent, the dorsal hippocampus may be more important for fine-grained spatial processing and the ventral hippocampus for what they call “context coding” (i.e., information about what a particular locale is like). In a final inferential step, they review fMRI studies of memory retrieval in the human and propose that “Retrieving detailed spatial relational information preferentially activate[s] the posterior hippocampus, whereas retrieving information about locales (or contexts) preferentially activate[s] the anterior hippocampus” (p. 22).

(Recall that another perspective on how the hippocampus may support episodic memory is summarized in *Web Clip 6.2*.)

WHAT ARE THE COGNITIVE FUNCTIONS OF THE HIPPOCAMPUS?

Some of the ideas reviewed in the two previous sections, like the discovery of Hebbian plasticity in the hippocampus, and the dependence of performance of

(at least some) tests of LTM on the NMDA receptor, fit well with the provisional principles that we listed in the earlier subsection *The organization of memory*. Others, like the putative role for the hippocampus in the retrieval of memories, would seem incompatible with the principles derived from initial studies of H. M. In this final section, we’ll revisit those principles articulated near the beginning of this chapter in light of more recent theoretical and empirical developments.

Standard anatomical model

After some puzzling early failures to replicate H. M.’s amnesia in nonhuman primates, by the 1980s, teams led by Mortimer Mishkin (yes, *that* Mort Mishkin) and by Larry Squire and Stuart Zola (then Zola-Morgan) produced a series of findings indicating that surgical lesioning of the amygdala and hippocampus, bilaterally, produced a marked impairment of LTM. The standard task was the delayed nonmatch-to-sample (DNMS), and bilateral MTL animals began to show significant impairments with delays between sample and test stimuli as short as 15 seconds. Notably, lesions targeting just selected portions of the MTL seemed to produce quantitatively, but not qualitatively, different patterns of impairment. Thus, for example, Squire (1992) summarizes that lesions restricted to the amygdala had no effect compared to unoperated controls, lesions of hippocampus plus the overlying parahippocampal gyrus (PHG) and caudal PRC (“H⁺”) produced a marked impairment at 15 seconds, and lesions of hippocampus plus parahippocampal cortex plus PRC and EC (“H⁺⁺”) produced an even larger impairment at 15 seconds. When the delay was extended to 60 seconds, then to 10 minutes, the rate of decline for each group (including controls) was comparable. Note that, with the surgical aspiration methods that were used to produce these lesions, a “H” lesion that did not compromise any of the overlying tissue was not possible, due to its location deep in the MTL.

The model that emerged from results such as these was of the hierarchy as drawn in *Figure 5.5.B*, with regions feeding into PRC and PHG classified as “perceptual,” and PHG, PRC, EC, and the hippocampus all part of the “MTL memory system.”

Challenges to the standard anatomical model

Subsequent research using different, more recently developed lesioning methods and behavioral tasks, however, have called this hierarchical view into question.

Newer lesioning methods

With regard to lesions, Elisabeth Murray and colleagues at the NIH's Laboratory of Neuropsychology have used injections of ibotenic acid to produce more circumscribed lesions. Ibotenic acid is a glutamate agonist that, through a phenomenon called excitotoxicity, destroys neurons whose dendrites it comes into contact with. Importantly, it spares so-called "fibers of passage," i.e., axons traveling through the injection site, but that neither initiate nor terminate in it, because these fibers lack glutamate receptors. Mechanical lesions, in contrast, whether produced by surgical excision, heat coagulation, high electrical current, or aspiration, will destroy any fibers that pass through the lesioned area. With ibotenic acid injections, in contrast, Murray and colleagues could, for example, produce clean "H" lesions. Doing so produced results strikingly different from the earlier findings: H lesions produced no impairment on DNMS and other tests of LTM, whereas selective lesions of PRC and EC produced profound impairments.

Perceptually challenging behavioral tasks

Research by Murray, Oxford-based colleague David Gaffan, and collaborators has also evaluated MTL function from the perspective that PRC, being one synapse downstream from AIT, may subserve high-level object recognition functions in addition to LTM encoding functions. In addition to the anatomy, they were motivated by computational modeling suggesting that PRC might be particularly important for representing complex conjunctions of features. *Figure 10.7* illustrates one experiment in which monkeys first had to learn, through trial-and-error guessing, which one in each pair of stimuli was rewarded (*Tip 10.6*). For the test, these stimulus pairs were re-presented as they had appeared during training, along with two additional pairs per training pair that were constructed by partially morphing the two trained images together. Thus, as illustrated in panel

A of *Figure 10.7*, the "feature ambiguity" pairs required discriminating between two stimuli that shared many features. The finding was that H- and PRC-lesioned animals did not differ when discriminating between trained stimuli, but the PRC-lesioned animals were selectively impaired on pairs with high feature ambiguity.

Importantly, Murray, Bussey, and Saksida (2007) note that PRC-lesioned animals are not impaired on just any difficult task. For example, they perform normally on tasks requiring difficult color or size discriminations. Thus, on the strength of these and other results, Murray and colleagues propose that "the perirhinal cortex participates in both perception and memory: it is important for perception because it contains mechanisms enabling the representation of complex conjunctions of features; it subserves memory because it serves as a storage site of complex stimulus representations in both the short- and long-term" (pp. 104–105). (*Web Clip 10.4* provides further elaboration of these ideas.)

The two previous subsections have summarized arguments that (1) the hippocampus is not needed for relatively simple object memory performance, whereas PRC and EC are; and (2) PRC performs important perceptual, in addition to mnemonic, functions. Both of these arguments represent serious challenges to the notion that the MTL can be considered a monolithic "memory system." Argument (1), in particular, may seem alarming, because it challenges the longstanding assumption that the hippocampus is critical for declarative memory encoding, full stop. The next subsection summarizes some more recent proposals for how to conceive of the role of the hippocampus in cognition.

Tip 10.6

This "concurrent discrimination" task, formally similar to the control task included in *Research Spotlight 10.2*, is thought to depend on plasticity in the neostriatum. This might explain normal learning on it by AP5-treated rats in Morris et al. (1986) and by H- and PRC-lesioned monkeys in Murray, Bussey, and Saksida (2007).

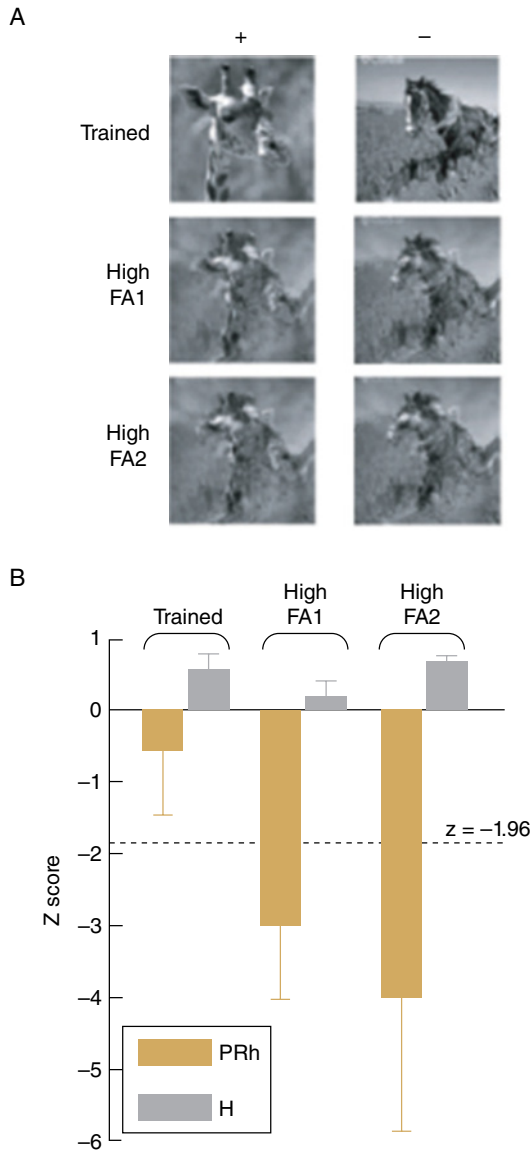


FIGURE 10.7 A. On each trial the monkey saw a pair of stimuli and had to select one. During training the same images were always paired, and the reward contingencies were always the same. At test, the animals saw each of the training pairs, along with two variants of each pair that contained two levels of “feature ambiguity” (*High FA1* and *High FA2*). **B.** Performance of animals with H vs. PRC (here, labeled “PRh”) lesions on the three levels of perceptual difficulty. Source: Murray, Bussey, and Saksida, 2007. Reproduced with permission of Annual Reviews.

“New roles” for the hippocampus?

One proposed reconceptualization of the hippocampus, championed by neurobiologist Howard Eichenbaum (from *Web Clip 6.2*) and cognitive neuroscientist Neal Cohen, is that its primary function is one of *relational binding*. In brief, this theory holds that the hippocampus effects the operation of representing and learning the relations between items in the environment. This might include the arbitrary rule for written English of “*i* before *e*, except after *c*,” or the concrete spatial content of something like “Zidane struck the free kick from the left side of the field, lofting the ball over the heads of the Brazilian defenders and into the right side of the goal box, where Henry, running in unmarked, volleyed it into the back of the net.” One implication of this idea is that the classical assumption that the hippocampus is not critical for STM may also need to be revisited. Indeed, recent studies have demonstrated that patients with hippocampal damage can be impaired on tests of STM for the spatial relations between items in a display, even with delays as brief as one second. Thus, from the relational binding perspective, the question isn’t *How long must the information be held in memory?* but *What kind of information needs to be processed?*

A second, relatively recent, conceptualization of the function of the hippocampus is a role in what some call “mental time travel,” the ability to anticipate a future event and imagine various outcomes, as well as the ability to imagine entirely made-up events. In one study, led by (Ireland-born, London-based) cognitive neuroscientist Eleanor Maguire, patients with bilateral hippocampal damage were asked to construct new imagined experiences, such as “Imagine you are lying on a white sandy beach in a beautiful tropical bay,” and “Imagine that you are sitting in the main hall of a museum containing many exhibits.” Their results indicated that the imagined experiences of the patients contained markedly less experiential richness than did those of healthy control subjects. A more detailed analysis also revealed lower “spatial coherence” (a measure of the contiguousness and spatial integrity of an imagined scene) in the performance of the patients. If the latter were at the root of the overall poor performance of the patients, these findings would be consistent with the hippocampus-as-cognitive-map framework.

A third development that postdates what we've called the standard model, that the hippocampus may be particularly critical for recollection-based, but not familiarity-based memory retrieval, will be considered in detail in *Chapter 11*.

Consolidation

The concept of consolidation has also become more complicated since the 1960s and 70s. Consistent with the original idea of a process that yielded a time-limited role for the hippocampus in LTM, we have reviewed that McClelland, McNaughton, and O'Reilly (1995) proposed a computational basis for consolidation as the gradual integration of hippocampally based representations into the neocortex. A theoretical challenge to the idea of consolidation, however, has come in the form of the *multiple trace theory*, proposed by Lynn Nadel and neuropsychologist Morris Moscovitch. The multiple trace theory posits that each instance of memory retrieval also prompts the creation by the hippocampus of a new memory trace, such that over time a single memory comes to be stored as multiple traces. To the extent that elements of these traces overlap, this process leads to the development of semantic knowledge that is independent of the episodes in which the information was learned. So, for example, if learning about US presidents in primary school and taking a family trip to Washington DC both create traces representing the proposition that *Thomas Jefferson was the third President of the United States*, repeated iterations of this process create a representation that can be retrieved independently of any reference to any one of the contexts in which this information was encountered. In this way, the memory that *Thomas Jefferson was the third President of the United States* becomes a *semantic* memory. Should damage to the hippocampus be sustained several years after the learning took place, the patient would nonetheless be able to retrieve this knowledge. On this prediction, the multiple trace theory and the standard neuropsychological model are in accord. The two differ, however, in that the former holds that the specific memory of the visit to the Jefferson Memorial during the family trip to Washington DC remains an *autobiographical episodic* memory that is dependent on the hippocampus for the remainder of the subject's life. Thus, multiple trace

Tip 10.7

This kind of "one-trial learning" is understood to be an important evolutionary adaptation: after avoiding a near scrape with death, it is the animals that never repeated the same mistake who survived to pass along their genetic material. A similar phenomenon, the "Garcia effect," is seen when one becomes sick after eating a new food, or a distinctive food, in a novel environment. Patients about to undergo chemotherapy are sometimes given an unusual food, such as a root beer-flavored piece of candy, so that they will develop an aversion to this, rather to whatever "normal" food they may have most recently eaten prior to the treatment.

theory would predict that access to this autobiographical episodic memory would be severely impoverished, if not completely impossible, after extensive damage to the MTL (particularly to the hippocampus). The standard neuropsychological model, in contrast, would hold that access to remote autobiographical episodic memories would be intact. This remains an area of vigorous and contentious debate.

Another puzzle related to consolidation – the phenomenon of reconsolidation and its potential relevance for the treatment of psychological trauma – is considered in *Real-World Applications 10.1*.

TO CONSOLIDATE

Memory is of central importance to human consciousness and cognition. At the conscious level, it creates our sense of identity, and provides the soothing, subjectively continuous internal narrative that enables us to go about our business without constantly stopping to ask "Why am I here? What was I doing 5 minutes ago?" At an unconscious level, it shapes our habits, our proclivities and dispositions, and makes us who we are. Thus, it's not surprising that different facets of memory are the focus of such intense research activity. Nor should it be a surprise when, if a large number of really smart, really driven people are working on a similar problem, multiple answers will emerge, not all of them mutually compatible.

REAL-WORLD APPLICATIONS

10.1 Reconsolidation

The phenomenon of reconsolidation was discovered in the context of a fear conditioning paradigm, a variant of Pavlovian conditioning in which an initially neutral stimulus (often a tone, or sometimes the novel experimental chamber in which the conditioning occurs (referred to as the “context”)) is paired with an aversive stimulus, often an electric shock delivered through the floor. It differs from the example with which we started off this chapter in that a single trial is sufficient for the animal to learn the conditioned response, which is to freeze the next time it hears the tone (or, is placed back into the chamber; *Tip 10.7*). Although we won’t consider this in detail until *Chapter 17*, plasticity that is critical for the consolidation of a conditioned fear memory occurs in the amygdala. Evidence for this is that infusion of a protein synthesis inhibitor into the amygdala blocks the expression of the conditioned response when the animal is presented with the tone a day later. (Without the protein synthesis inhibitor, control animals freeze when they again hear the tone, and the duration of freezing is a dependent measure of how strong the memory is.)

Evidence for a reconsolidation process was produced by Karim Nader and colleagues in the laboratory of neurobiologist Joseph LeDoux when, instead of administering the protein synthesis inhibitor at the time of the learning episode, they administered it when the conditioned stimulus was re-presented, i.e., at the time of memory retrieval. The finding was that administering the protein synthesis inhibitor at the time of retrieval trial 1 did not disrupt the expression of the memory on retrieval trial 1; however, it did disrupt the expression of the memory on retrieval trial 2. This finding was controversial at first, but has since been replicated several times, in several species (including humans), and is now widely accepted. Although the mechanism(s) underlying reconsolidation aren’t fully understood, the implication is that the act of retrieving a memory puts it into a labile state that renders it vulnerable to disruption. (Here, it might be helpful to think of, e.g., Lisman’s idea for how the activation of one item in a sequence can then trigger the activation of the subsequent items.) Part of the puzzle is that the LTP underlying the memory has already been through the (or, at least, a) structural phase. Is it reversed? Or somehow overwritten?

Even though the neurobiological and computational details of how and why reconsolidation happens remain uncertain, the finding of Nader, Schafe, and LeDoux (2000) very quickly initiated considerable excitement about potential clinical applications. As Schwabe and colleagues (2012) note: “Emotionally arousing experiences are usually better remembered than neutral experiences. Although generally adaptive to survival, this emotional memory enhancement may contribute to anxiety disorders, such as post-traumatic stress disorder (PTSD)” (p. 380). Therefore, could one leverage the phenomenon of reconsolidation to develop a “pill to forget” for treating victims of trauma? The study summarized in *Figure 10.8* represents an initial step in this direction, by demonstrating that a pharmacological intervention delivered at the time of retrieval of an emotional memory can, in effect, “reduce its emotionality” for subsequent retrieval episodes.

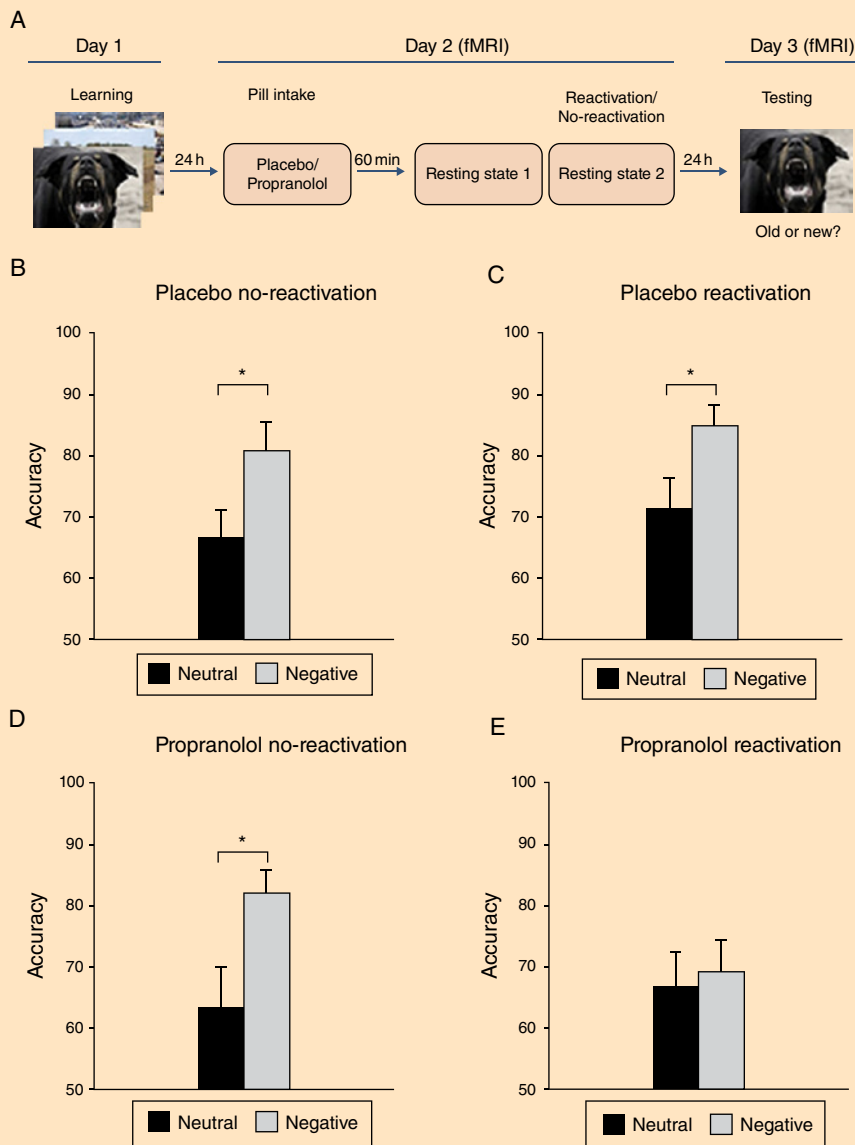
Real-World Applications 10.1 *Continued*

FIGURE 10.8 Weakening a consolidated emotional memory by disrupting reconsolidation. On Day 1 subjects viewed 25 emotionally negative and 25 emotionally neutral images. On Day 2 they either did or didn't receive a β -adrenergic antagonist ("beta-blocker") ("Propranolol"/"Placebo") and either were or weren't tested on their memory for the images ("Reactivation"/"No-reactivation"). (fMRI data were also acquired, but are not considered here.) On Day 3, memory for the images was tested in each of the four groups, and memory for emotionally negative images was impaired only in the "Propranolol reactivation" group. (In *Chapter 17*, we'll consider the role of norepinephrine in emotional processing.) Source: Schwabe, Nader, Wolf, Beaudry, and Pruessner, 2012. Reproduced with permission of Elsevier.

The next three chapters will build on the themes that we have introduced here. Not all of the controversies will be resolved, but the reader will be equipped with a

good sense of what some of the most interesting, most important problems are, and the ways that some of the leading scientists in the field are addressing them.

END-OF-CHAPTER QUESTIONS

1. What aspect of the global amnesic syndrome suggests a distinction between LTM encoding and LTM retrieval? Why?
2. What aspect of the global amnesic syndrome suggests a process of consolidation in LTM formation?
3. State Hebb's postulate. (Paraphrasing is okay, it's the idea that's important.)
4. In what sense does the NMDA receptor act as a coincidence detector? How is this important for LTP?
5. What principle is emphasized in the model of McClelland, McNaughton, and O'Reilly (1995)? Of Lisman (1999)? Of Buszáki (2006)?
6. What is the principal difference between the standard model of consolidation and that of the multiple trace theory? What different predictions do the two make with respect to dependence on the hippocampus of autobiographical episodic memories?
7. Summarize the evidence for a process of reconsolidation.
8. In what respects are the experiments of Morris et al. (1986) and of Nader, Schafe, and Le Douarin (2000) similar, and in what respects are they different?

REFERENCES

- Bliss, Timothy Vivian Pelham, and Terje Lomo. 1973. "Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anesthetized Rabbit Following Stimulation of the Perforant Path." *Journal of Physiology* 232 (2): 331–341. <http://jp.physoc.org/content/232/2/331.full.pdf+html>.
- Buszáki, György. 2006. *Rhythms of the Brain*. Oxford: Oxford University Press.
- Corkin, Suzanne, David G. Amaral, R. Gilberto González, Keith A. Johnson, and Bradley T. Hyman. 1997. "H. M.'s Medial Temporal Lobe Lesion: Findings from Magnetic Resonance Imaging." *Journal of Neuroscience* 17 (10): 3964–3979. <http://www.jneurosci.org/content/17/10/3964.full.pdf+html>.
- Hebb, Donald Olding. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley and Sons.
- Hilts, Philip J. 1996. *Memory's Ghost: The Nature of Memory and the Strange Tale of Mr. M.* Touchstone edition. New York: Simon and Schuster. First published 1995.
- Lisman, John E. 1999. "Relating Hippocampal Circuitry to Function: Recall of Memory Sequences by Reciprocal Dentate–CA3 Interactions." *Neuron* 22: 233–242.
- Lomo, Terje. 2003. "The Discovery of Long-Term Potentiation." *Philosophical Transactions of The Royal Society of London, Series B: Biological Sciences* 358 (1432): 617–620. doi: 10.1098/rstb.2002.1226.
- Lüscher, Christian, and Robert C. Malenka. 2012. "NMDA Receptor-Dependent Long-Term Potentiation and Long-Term Depression (LTP/LTD)." *Cold Spring Harbor Perspectives in Biology* 4 (6): a005710. doi: 10.1101/cshperspect.a005710.
- McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly. 1995. "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory." *Psychological Review* 102 (3): 419–457. doi: 10.1037/0033-295X.102.3.419.
- Morris, Richard G. M., E. Anderson, Gary S. Lynch, and Michael Baudry. 1986. "Selective Impairment of Learning and Blockade of Long-Term Potentiation by an N-Methyl-D-Aspartate Receptor Antagonist, AP5." *Nature* 319 (6056): 774–776. doi: 10.1038/319774a0.
- Murray, Elisabeth A., Timothy J. Bussey, and Lisa M. Saksida. 2007. "Visual Perception and Memory: A New View of Medial Temporal Lobe Function in Primates and Rodents." *Annual Review of Neuroscience* 30: 99–122. doi: 10.1146/annurev.neuro.29.051605.113046.
- Nadel, Lynn, Siobhan Hoscheidt, and Lee R. Ryan. 2013. "Spatial Cognition and the Hippocampus: The Anterior–Posterior Axis." *Journal of Cognitive Neuroscience* 25 (1): 22–28. doi: 10.1162/jocn_a_00313.

- Nader, Karim, Glenn E. Schafe, and Joseph E. Le Doux. 2000. "Fear Memories Require Protein Synthesis in the Amygdala for Reconsolidation after Retrieval." *Nature* 406 (6797): 722–726. doi: 10.1038/35021052.
- O'Keefe, John, and Lynn Nadel. 1978. *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press.
- Pohl, Walter. 1973. "Dissociation of Spatial Discrimination Deficits Following Frontal and Parietal Lesions in Monkeys." *Journal of Comparative and Physiological Psychology* 82 (2): 227–239. doi: 10.1037/h0033922.
- Scoville, William Beecher, and Brenda Milner. 1957. "Loss of Recent Memory after Bilateral Hippocampal Lesions." *Journal of Neurology, Neurosurgery, and Psychiatry* 20 (1): 11–21. doi: 10.1136/jnnp.20.1.11.
- Schwabe, Lars, Karim Nader, Oliver T. Wolf, Thomas Beaudry, and Jens C. Pruessner. 2012. "Neural Signature of Reconsolidation Impairments by Propranolol in Humans." *Biological Psychiatry* 71 (4): 380–386. doi: 10.1016/j.biopsych.2011.10.028.
- Squire, Larry R. 1992. "Memory and the Hippocampus: A Synthesis from Findings with Rats, Monkeys, and Humans." *Psychological Review* 99 (2): 195–231. doi: 10.1037/0033-295X.99.3.582.

OTHER SOURCES USED

- Baddeley, Alan D., and Elisabeth K. Warrington. 1970. "Amnesia and the Distinction between Long- and Short-Term Memory." *Journal of Verbal Learning and Verbal Behavior* 9 (2): 176–189. doi: 10.1016/S0022-5371(70)80048-2.
- Cooper, Jack R., Floyd E. Bloom, and Robert H. Roth. 1991. *The Biochemical Basis of Neuropharmacology*, 6th ed. New York: Oxford University Press.
- Corkin, Suzanne. 2013. *Permanent Present Tense: The Unforgettable Life of the Amnesic Patient, H. M.* New York: Basic Books.
- Eichenbaum, Howard, and Neal J. Cohen. 2001. *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. Oxford Psychology Series. Oxford: Oxford University Press.
- Fuller, Leah, and Michael E. Dailey. 2007. "Preparation of Rodent Hippocampal Slice Cultures." *Cold Spring Harbor Protocols* 2007 (10): 4848–4852. doi: 10.1101/pdb.prot4848.
- Gabrieli, John D. E., M. M. Keane, Ben Z. Stanger, Margaret Kjelgaard, Suzanne Corkin, and John H. Growdon. 1994. "Dissociations Among Structural-Perceptual, Lexical-Semantic, and Event-Fact Memory in Amnesia, Alzheimer's, and Normal Subjects." *Cortex* 30: 75–103.
- Jarrard, Leonard E. 1989. "On the Use of Ibotenic Acid to Lesion Selectively Different Components of the Hippocampal Formation." *Journal of Neuroscience Methods* 29 (3): 251–259. doi: 10.1016/0165-0270(89)90149-0.
- Milner, Brenda, Suzanne Corkin, and H.-L. Teuber. 1968. "Further Analysis of the Hippocampal Amnesic Syndrome: 14-Year Follow-Up Study of H. M." *Neuropsychologia* 6 (3): 215–234. doi: 10.1016/0028-3932(68)90021-3.
- Nadel, Lynn, and Morris Moscovitch. 1997. "Memory Consolidation, Retrograde Amnesia and the Hippocampal Complex." *Current Opinion in Neurobiology* 7 (2): 217–227. doi: 10.1016/S0959-4388(97)80010-4.
- Nader, Karim, and Oliver Hardt. 2009. "A Single Standard for Memory: The Case for Reconsolidation." *Nature Reviews Neuroscience* 10 (3): 224–234. doi: 10.1038/nrn2590.
- Pearce, John M. S. 2001. "Ammon's Horn and the Hippocampus." *Journal of Neurology, Neurosurgery, and Psychiatry* 71 (3): 351. doi: 10.1136/jnnp.71.3.351.
- Schwabe, Lars, Karim Nader, and Jens C. Pruessner. 2013. "β-Adrenergic Blockade during Reactivation Reduces the Subjective Feeling of Remembering Associated with Emotional Episodic Memories." *Biological Psychiatry* 92 (2): 227–232. doi: 10.1016/j.biopsych.2012.10.003.
- Squire, Larry R., and Stuart Zola-Morgan. 1991. "The Medial Temporal Lobe Memory System." *Science* 253 (5026): 1380–1386. doi: 10.1126/science.1896849.
- Squire, Larry R., Craig E. L. Stark, and Robert E. Clark. 2004. "The Medial Temporal Lobe." *Annual Review of Neuroscience* 27: 279–306. doi: 10.1146/annurev.neuro.27.070203.144130.

FURTHER READING

- Clelland, Claire D., M. Choi, C. Romberg, G. D. Clemenson Jr., Alexandra Fragniere, P. Tyers, ... Tim J. Bussey. 2009. "A Functional Role for Adult Hippocampal Neurogenesis in Spatial Pattern Separation." *Science* 325 (5937): 210–213. doi: 10.1126/science.1173215.
- The precise role(s) of neurogenesis – the "birth" of new neurons and their migration into the dentate gyrus – in LTM are still being worked out, but this paper captures much of the excitement and potential importance of the phenomenon.*

- Corkin, Suzanne. 2002. "What's New with the Amnesic Patient H. M.?" *Nature Reviews Neuroscience* 3 (2): 153–160. doi: 10.1038/nrn726.
From a student of Brenda Milner who has studied H. M. intensively from the 1960s up through the final years of his life.
- Corkin, Suzanne. 2013. *Permanent Present Tense: The Unforgettable Life of the Amnesic Patient, H. M.* New York: Basic Books.
A book that with equal parts compassion and scientific acuity, alternates between the biography and tragedy of Henry Molaison, and the manifold discoveries about human memory that have resulted from research with this remarkable patient.
- Dębiec, Jacek. 2012. "Memory Reconsolidation Processes and Posttraumatic Stress Disorder: Promises and Challenges of Translational Research." *Biological Psychiatry* 71 (4): 284–285. doi: 10.1016/j.biopsych.2011.12.009.
Commentary addressing the implications of the study presented in Research Spotlight 10.1.
- Eichenbaum, Howard. 2013. "What H. M. Taught Us." *Journal of Cognitive Neuroscience* 25 (1): 14–21. doi: 10.1162/jocn_a_00285.
Review from a leading expert on LTM and the functions of the hippocampus.
- Frankland, Paul W., Stefan Köhler, and Sheena A. Josselyn. 2013. "Hippocampal Neurogenesis and Forgetting." *Trends in Neuroscience* 36 (9): 497–503. doi: 10.1016/j.tins.2013.05.002.
MTL neurogenesis is highest early in life, and this review suggests that this may account for the phenomenon of "infantile amnesia" – our inability to remember events from the first few years of our lives.
- Gaffan, David. 2002. "Against Memory Systems." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 357 (1424): 1111–1121. doi: 10.1098/rstb.2002.1110.
A provocative set of arguments for a radical rethinking of LTM. I don't agree with all of it, but nonetheless always assign it to my grad seminar on the Cognitive Neuroscience of Memory.
- Luu, Paul, Orriana C. Sill, Lulu Gao, Suzanna Becker, Jan Martin Wojtowicz, and David M. Smith. 2012. "The Role of Adult Hippocampal Neurogenesis in Reducing Interference." *Behavioral Neuroscience* 126 (3): 381–391. doi: 10.1037/a0028252.
Whereas Frankland, Köhler, and Josselyn (2013) propose that MTL neurogenesis may be responsible for retroactive interference (in the very young), this paper presents a scheme whereby MTL neurogenesis in the adult may aid in the individuation of distinct episodic memories.
- Nader, Karim, and Oliver Hardt. 2009. "A Single Standard for Memory: The Case for Reconsolidation." *Nature Reviews Neuroscience* 10 (3): 224–234. doi: 10.1038/nrn2590.
A vigorous defense of the phenomenon of reconsolidation that, in the process, provides a nice review of the principles of consolidation and reconsolidation in fear conditioning.
- Oudiette, Delphine, and Ken A. Paller. 2013. "Upgrading the Sleeping Brain with Targeted Memory Reactivation." *Trends in Cognitive Sciences* 17 (3): 142–149. doi: 10.1016/j.tics.2013.01.006.
A role for sleep in memory consolidation is revealed when LTM for specific items can be improved when auditory or olfactory stimuli that are associated with those items are presented to subjects while they sleep.
- Postle, Bradley R. 2009. "The Hippocampus, Memory, and Consciousness." In *Neurology of Consciousness*, edited by Steven Laureys and Giulio Tononi, 326–338. Amsterdam: Elsevier.
A review chapter by Yours Truly that goes into more detail about the pathophysiology of amnesia and the cognitive neuroscience of long-term memory.

CHAPTER 11

DECLARATIVE LONG-TERM MEMORY

KEY THEMES

- Early neuroimaging studies of LTM encoding were difficult to design, and their results difficult to interpret, due to uncertainty over how to activate the hippocampus differentially in different conditions, and how to unconfound effects of novelty from effects of encoding per se.
- fMRI studies of the subsequent memory effect, made possible by the development of “fast” event-related techniques, have identified PFC and inferior temporal regions associated with level-of-processing manipulations as being important for successful memory encoding.
- Another important innovation in neuroimaging data analysis, particularly for studies of LTM, is methods for measuring functional connectivity between brain regions.
- Intensive memorization of a highly detailed and idiosyncratic 2-D map, followed by daily navigation through this environment (i.e., being a licensed taxi driver in London), leads to gross anatomical expansion of the posterior hippocampus, and commensurate reduction of the anterior hippocampus.
- Accurate interpretation of impaired vs. spared performance of anterograde amnesic patients on tests of LTM requires an understanding of the level of awareness of the learning episode that is required by the task.
- MVPA of fMRI data can be used to predict the category of information that a subject is about to recall, a feat that wouldn’t be possible were contextual reinstatement not a principle underlying recall from LTM.
- Models from cognitive psychology that posit a distinction between processes of familiarity vs. recollection in memory retrieval have received considerable support from cognitive neuroscience research, with the latter being explicitly linked to the hippocampus.

CONTENTS

KEY THEMES

THE COGNITIVE NEUROSCIENCE OF LTM ENCODING

Neuroimaging the hippocampus

Novelty-driven “encoding”

*Subsequent memory effects in the PFC and
MTL*

*Uneven evidence for encoding-related activity
in hippocampus*

Incidental encoding into LTM during a short-term
memory task

*Signal intensity-based analysis of the subse-
quent memory effect*

*A short-term memory buffer to bridge the
temporal gap between to-be-associated
events*

*Functional connectivity-based analysis of the
subsequent memory effect*

THE HIPPOCAMPUS IN SPATIAL MEMORY EXPERTS

RETRIEVAL

Retrieval without awareness

Documenting contextual reinstatement in
the brain

*The dynamics of multivariate state changes
underlying behavior*

*Implications for “mind reading” ... well, “brain
reading”*

Familiarity vs. recollection

Dual-process models

*Dual processes of memory retrieval in the
rodent*

KNOWLEDGE

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

THE COGNITIVE NEUROSCIENCE OF LTM

The previous chapter emphasized the neurobiology of memory. With that groundwork in place, this chapter will focus much more on cognitive-level questions, in many instances looking at how longstanding cognitive theories (e.g., levels of processing, contextual reinstatement) have been studied by cognitive neuroscientists. So let's jump in!

ENCODING

Neuroimaging the hippocampus

Neuroimaging studies of LTM from the 1980s and 1990s were particularly challenging to design, and their results challenging to interpret, due to the constraints imposed by block designs. To understand why, let's think back to how the Washington University group went about studying the visual perception of words (*Figure 6.5*). They presented words, serially, for a block of 40 second, and compared that with a 40 second block of fixation. Although one might (and some did) object that there was insufficient experimental control to support strong interpretation of what cognitive functions were engaged during the word blocks, there was little ambiguity about the control task: staring at an unchanging fixation point would produce minimal activation of the visual system relative to 1 Hz presentation of words. (And, indeed, the [words – fixation] subtraction revealed considerably higher levels of activity in ventral occipitotemporal cortex for the former than the latter.) It's not clear, however, that the same approach would work if one sought to “image the hippocampus,” primarily because it's believed that this structure is always encoding. For example, had subjects in the Petersen et al. (1988) study been debriefed an hour, or a day, after their scanning session, they would surely have remembered having endured the blocks of fixation as well as the blocks of word viewing. Thus, there's no reason to posit that their hippocampi would be any less (or more) active during the control than the experimental task. Were they differently active during the two conditions? Sure, but MVPA (and, therefore, any hope

of detecting such a difference with neuroimaging) was more than a decade away.

A variant of this problem will also crop up in the section of this chapter that focuses on retrieval from LTM, and is one that was foreshadowed in the previous chapter's section on reconsolidation: even the act of retrieving information from LTM will be encoded by the hippocampus. If the darned thing never shuts off, how can we measure its function with a subtractive methodology?

Novelty-driven “encoding”

The reasoning laid out in the previous section was invoked often in the early to mid-1990s, when a series of PET studies of encoding into LTM failed to show evidence for differential hippocampal activity during encoding vs. control blocks. One way that experimenters tried to address this problem was to reason, in effect, that *if we can't devise a condition when it is not active, perhaps we can at least devise one when it is less active*. To achieve this, two groups, one led by (Estonia-born, Toronto-based) cognitive psychologist Endel Tulving, another by Chantal Stern, Suzanne Corkin, and colleagues in Boston, explicitly controlled novelty. Tulving, Markowitsch, Kapur, Habib, and Houle (1994) pre-exposed subjects to 40 color photos of complex scenes the day prior to the scanning session, then scanned subjects on blocks that either featured the “old” items from the day before, or 40 new items that they had never seen before. Although one can imagine that the “old” condition might have generated a considerable amount of memory retrieval, what Tulving et al. (1994) emphasized was that the “new” items had greater novelty, and thus were likely to emphasize encoding processes. Their findings included greater “new” than “old” activity in the hippocampus and parahippocampal gyrus.

Stern et al. (1996) also presented 40 novel images during the experimental block, and instructed subjects to encode them for a subsequent memory test. During the control block they presented a single image (also novel) during the entire duration of the block. Similar to the Tulving et al. (1994) findings, this study also found greater activity for the varying-identity stimuli than the repeated-identity stimulus in the posterior hippocampus and parahippocampal gyrus, as well as more posteriorly in the lingual and fusiform gyri.

In retrospect, knowing what we do now about, for example, category-preferential responses in the ventral occipitotemporal cortex (*Chapter 9*), we can surmise that at least some of the activity reported in these studies related more to visual processing than to memory encoding per se. Additionally, both explicitly confounded stimulus novelty with the mental process that they sought to engage. Thus, even at the time that these studies appeared, more impactful than their conceptual contributions was the fact that they, and a handful of other contemporaneous studies, succeeded in “activating” the elusive hippocampus. This had the effect of accelerating the pace at which the memory research community worked to develop better methods that would produce more definitive results. Indeed, two years after the Stern et al. (1996) study, the laboratory of a former student of Corkin, and a group of associates of Corkin and Stern at the Massachusetts General Hospital (MGH) NMR Center, simultaneously published breakthrough studies of memory encoding using novel event-related techniques.

Subsequent memory effects in the PFC and MTL

Since the early 1980s, ERP researchers had employed the following procedure to study encoding into LTM: (1) present subjects with several items, one at a time; (2) create a series of ERPs time-locked to stimulus onset; (3) after the EEG recording session, administer a surprise memory test for the items that had been presented during the EEG session; (4) selectively average the ERPs according to whether each item was or was not remembered in the subsequent memory test. Any differences between the two resultant ERPs would identify signal associated with successful vs. unsuccessful memory encoding (*Tip 11.1*). The logic

of this approach has much to recommend it. In particular, unlike with the neuroimaging studies reviewed in the previous section, subjects are performing the same task on every trial, and so this avoids the problem of whether differences in physiological signal are due to the memory-encoding processes of interest, or to some other factor (novelty per se, boredom, level of vigilance, etc.). The problem with the effect in the ERP literature was that it was capricious – not every lab could produce it, and even within a lab the effect would come and go depending on such factors as whether the subsequent memory test was a test of free recall, cued recall, or recognition. The latter was particularly puzzling because the ERP signal, of course, was collected prior to the time that the subsequent memory test would be administered. If subjects didn’t even know that there was going to be a memory test, how could it be that the format of that future test would influence encoding-related activity? [*Hey, that’d be a good extra-credit question ;-).*] The result of these problems was that interest in this approach had largely fizzled out. Until, that is, those fMRI people came along ...

Randy Buckner earned his PhD at Washington University in St. Louis, under the tutelage of Petersen, Raichle, and colleagues. On the strength of his graduate training in PET imaging he was invited to the MGH NMR Center as a kind of postdoctoral “Minister without Portfolio” (*Tip 11.2*). He gravitated to Cambridge, MA, memory luminaries Daniel Schacter and Suzanne Corkin, and, together with Schacter, postdoctoral fellow Anthony Wagner, and others, implemented the subsequent memory paradigm with fMRI. At the same time, back at the Stanford University laboratory where Wagner had recently earned his PhD, former Corkin student John Gabrieli and *his* student James Brewer were carrying out a very similar study. The two were published, back-to-back in *Science*, in 1998, and the Wagner et al. (1998) study is detailed in *Research Spotlight 11.1*.

It is important to note that neither the Wagner nor the Brewer study identified a “memory network,” per se. Wagner et al. (1998), in their study of memory for words, identified exclusively left hemisphere regions. Brewer, Zhao, Desmond, Glover, and Gabrieli (1998), although using a similar subsequent-memory design and analysis, used photos of indoor and outdoor

Tip 11.1

This procedure assays so-called “incidental encoding,” in that there were no explicit instructions to commit items to memory. Of course, > 99% of the things that we remember from our past are products of incidental encoding.

RESEARCH SPOTLIGHT

11.1 Successful encoding into LTM

Wagner and colleagues (1998) designed their study within the levels-of-processing framework, the venerable finding from cognitive psychology that a stimulus is better remembered when one thinks about its meaning (processes it at a “deep” level) than when one thinks about a superficial property (processes it at a “shallow” level). Their first experiment did this explicitly, presenting individual words to subjects and instructing them to either judge whether each corresponded to a living or nonliving thing (deep processing) or, during alternate blocks, whether each was printed in upper- or lowercase type. A surprise memory test after the fMRI scanning session confirmed that the manipulation worked, with subjects subsequently recognizing 85% of deeply processed items vs. 47% of shallowly processed items. In the brain, regions showing greater activity on deep vs. shallow blocks included left inferior frontal gyrus (IFG) of the ventrolateral PFC, and left parahippocampal and fusiform gyri on the ventral surface of the temporal lobe. This set up predictions for the subsequent memory study: Finding regions of overlap would be consistent with the idea that item-to-item fluctuations in how we think about information determines, at least in part, how effectively we encode it into LTM.

For the subsequent memory study, fMRI data were acquired while subjects viewed a total of 480 words across six blocks. Each block included 40 abstract words, 40 concrete words, and 40 fixation events, presented in a randomly determined order for 2 seconds each. This presentation scheme allowed for a “fast” event-related analysis with a method that Buckner had recently developed with then-MGH NMR Center colleague Anders Dale. They referred to the method as “selective averaging,” a label that captures some of the intuition about how it works, but not all the detail. Think back to *Figure 6.8*, which illustrates how even the briefest of events produces a “sluggish” response that doesn’t peak until 4–6 seconds after the neural event of interest, and doesn’t fully return to baseline until 20 or more seconds. This means that Wagner et al. (1998) would have expected variance attributable to the first item of a trial to still be influencing the fMRI signal when the tenth item was being presented, not to mention variance from the other eight intervening items. A short-hand label that comes closer to what the Dale and Buckner (1997) method accomplishes is “linear deconvolution” – using analytic methods to partition variance among trial types and thereby construct an estimate of the isolated hemodynamic response to each (*Tip 11.3*).

Although each scanning session presented 240 concrete and 240 abstract words, the data weren’t analyzed in this way. Rather, they were analyzed according to whether or not subjects remembered them in the surprise, post-scan recognition test: They saw the 480 scanned words and 480 new words and indicated for each whether they were highly confident that it was an “old” word, less confident that it was old, or that it was new. *Figure 11.1* shows regions for which items subsequently recognized with “high-confidence hits” generated more activity than did items that were subsequently forgotten (i.e., “misses”). Note that this effect occurred in essentially all of the regions that had been identified in the levels-of-processing experiment.

Research Spotlight 11.1 *Continued*

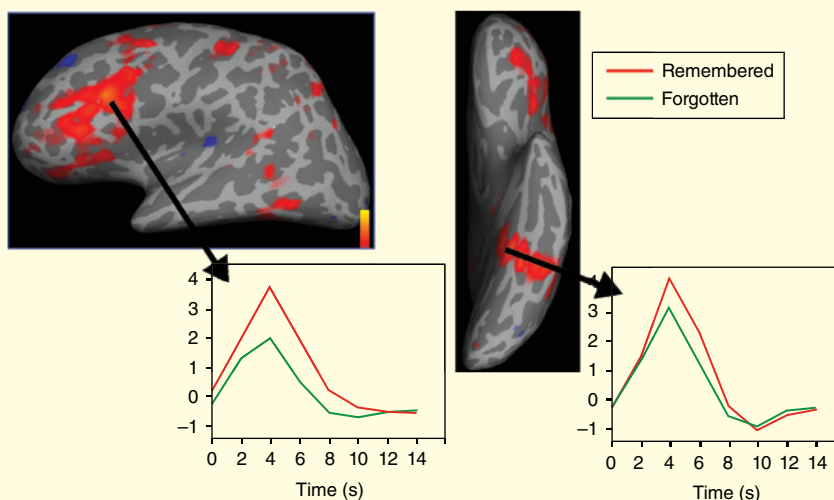


FIGURE 11.1 The subsequent memory effect reported in Wagner et al. (1998). Statistical maps, superimposed on two views of an inflated left hemisphere, show regions exhibiting higher activity for words that were later recognized with high confidence (“remembered”) than for words that were later not recognized. Inset plots show the fMRI evoked responses, estimated from the event-related analysis, from posterior IFG (left) and a region spanning parahippocampal and fusiform gyri (right; responses plotted in arbitrary units of BOLD signal intensity). Source: Adapted from Paller and Wagner, 2002. Reproduced with permission of Elsevier.

Tip 11.2

The MGH NMR Center (now the Martinos Center for Biomedical Imaging) has produced a remarkable amount of influential fMRI research, including the studies by Stern et al. (1996) and Kanwisher, McDermott, and Chun (1997) that we have previously considered.

Tip 11.3

This is referred to as a “fast” event-related design because to-be-separated events occur as quickly as once every 2 seconds; in “slow” event-related designs (*Methodology Box 8.1*), in contrast, event order can’t be randomized, and modeled events thus need to be separated by several seconds (*Web Clip 3.2*).

scenes, and the regions in which they found greater activity for high-confidence hits vs. misses were the parahippocampal gyrus, bilaterally, and right dorsolateral PFC. Thus, what both studies found was that activity in the very regions that are associated with verbal stimulus processing (in this case, reading words) and visual stimulus processing (looking at pictures),

respectively, are important determinants of whether or not that information will later be remembered. (We’ll see in *Chapter 18* that language processing is heavily biased to the left hemisphere. Additionally, other studies from Gabrieli’s group had indicated that the left IFG, in particular, is recruited for tasks

requiring semantic processing in the absence of any overt demands on episodic encoding or retrieval. And we've already seen that the PHG is important for representing information about "places.") The fact that these studies both identified the magnitude of activity in their respective networks to be an important factor in subsequent memory retrieval is consistent with the idea that the strength of attentional focus may be an important determinant in LTM encoding. They are also consistent with the idea, recurring throughout this book, that ongoing fluctuations in brain state can influence the quality of cognition at any moment in time.

Uneven evidence for encoding-related activity in hippocampus

It is noteworthy that neither of these two studies, the two most sophisticated fMRI studies of memory encoding up to that time, identified differential activity in the hippocampus during the processing of items that would subsequently be remembered vs. subsequently forgotten. Importantly, there are two conclusions that can NOT be drawn from this. At the physiological level, it would be a mistake to conclude that the hippocampus was somehow "not active" during these scans, for reasons that we've taken up previously. At the functional level, it would be a mistake to conclude from these studies, alone, that the hippocampus does not contribute to encoding into LTM. An interpretation from the perspective of the standard model, as outlined in the previous chapter, might be that the hippocampus is "doing its best" to encode everything that we experience, all the time, and that those thoughts and experiences that are somehow more salient, that are processed to a deeper level, are the ones that tend to get remembered. The primary function of the hippocampus, from this perspective, is to bind together the cortical representations whose association, after all, is what constitutes the memory. (Of course, "hippocampal revisionists" might point to the findings from Murray and colleagues that were also summarized in the previous chapter, and to the control condition from Morris, Anderson, Lynch, and Baudry (1986; *Research Spotlight 10.2*), to argue that hippocampal LTP isn't necessary for all forms of declarative memory, particularly for relatively simple stimulus recognition memory.)

Incidental encoding into LTM during a short-term memory task

Short-term memory (STM) will be the focus of *Chapter 13*, but one study of STM merits consideration here, for four reasons. First, it reinforces the idea that incidental encoding happens constantly, albeit with fluctuating efficacy, during the waking state (and even during some stages of sleep), regardless of what behavior we're engaged in. (The didactic point being that this is true *even* when we are engaged in a task that is nominally testing something other than LTM.) The second reason for considering this study is because it illustrates the power of relating individual differences in physiological signal to individual differences in behavior. Third, it relates to a mechanism that we considered in the previous chapter's section on how the hippocampus might work. Finally, it provides an opportunity to introduce another important method for the analysis of functional neuroimaging data.

Signal intensity-based analysis of the subsequent memory effect

Each trial in this study, from the laboratory of cognitive neuroscientist Charan Ranganath, began with the subject viewing a sample line drawing of a novel 3-D shape (multifaceted, blocky objects). After a delay period varying from 7 to 13 seconds in length, a memory probe appeared that matched the sample 50% of the time. After the scanning session, subjects were given a surprise test of recognition of the items that had been presented as sample stimuli during the STM task, and the fMRI data sorted as a function of subsequent LTM recognition. As illustrated in *Figure 11.2*, activity in the left dorsolateral PFC and the left hippocampus during the early delay period of the STM task predicted subsequent LTM for each item. This illustrates the point that, in principle, one should be able to find a subsequent memory effect for *any* kind of behavior, not just in experiments designed to assess LTM encoding.

Another notable feature of this study is illustrated in *Figure 11.2.B*. Here, the authors show that, on a subject-by-subject basis, the magnitude of each subject's neural subsequent-memory effect is related to his or her subsequent LTM for items that had been presented during the test of STM. Stated another

FIGURE 11.2 Subsequent LTM effects in left dIPFC and left hippocampus during a test of visual STM. Source: Ranganath, Cohen, and Brozinsky, 2005. Reproduced with permission of The MIT Press.

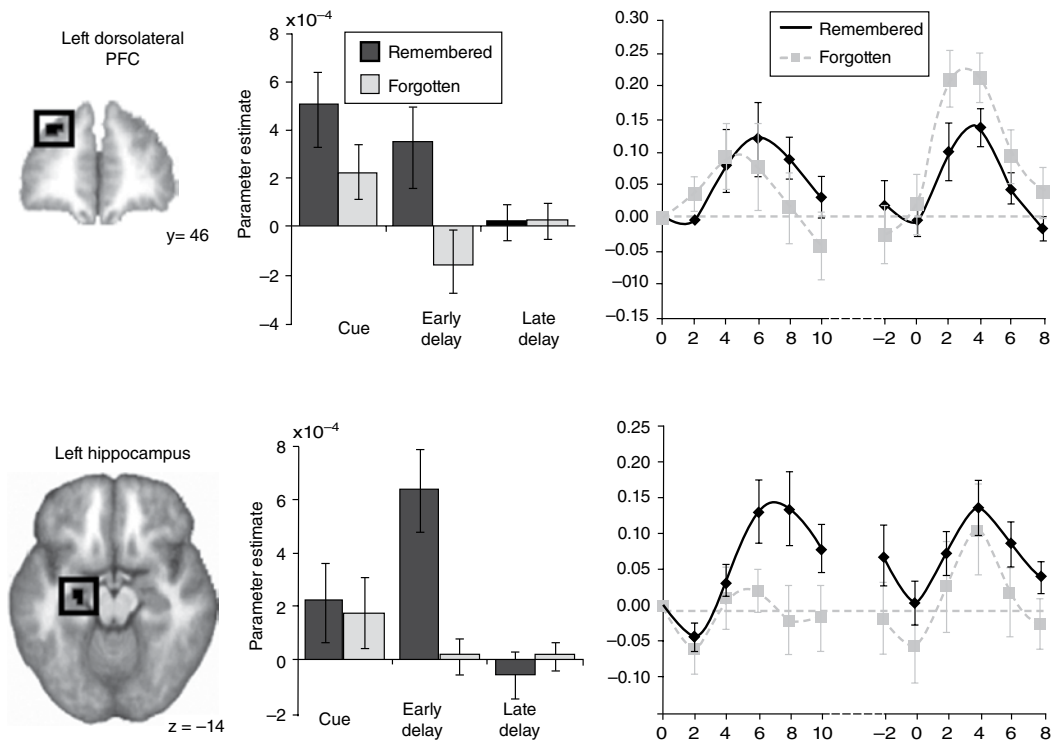


FIGURE 11.2.A Left-hand column: Regions of left dIPFC (coronal view) and left hippocampus (axial view) showing the effect. Center column: Bar graphs show magnitude of response during sample presentation ("Cue") and two portions of the delay period, expressed as parameter estimates from the GLM, as a function of whether they were subsequently remembered or forgotten on the surprise LTM test. Right-hand column: Time series show activity (% signal change from baseline as a function of time), time-locked to sample onset (first 0) and probe onset (second 0), also averaged as a function of subsequent LTM.

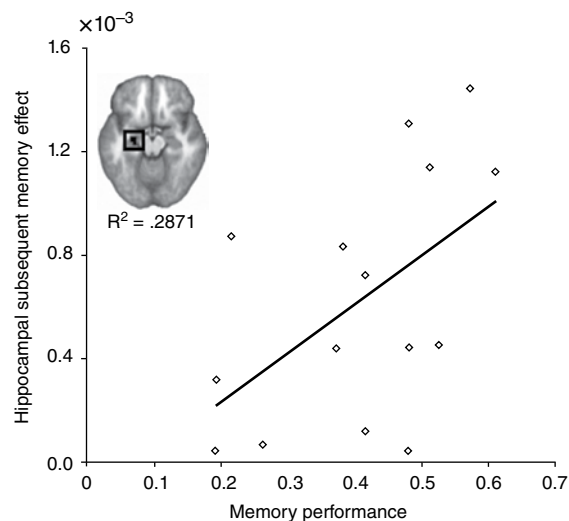


FIGURE 11.2.B Strength of early-delay hippocampal activity (vertical axis) predicts subsequent LTM performance.

way, the magnitude of the subsequent memory effect in left hippocampus predicted how well the subject would later remember items from the test of STM. Such a demonstration of **individual differences** can be powerful, because it suggests that the mechanism under study might explain, at least in part, why some individuals remembered more items on the subsequent memory test and why others remembered fewer. Note that this is still (indeed, explicitly) a test of correlation, and therefore doesn't establish a causal role for, in this case, hippocampal activity. However, one could argue that such a relationship would be a necessary (though not sufficient) property of this system if it were causally related to LTM formation.

A short-term memory buffer to bridge the temporal gap between to-be-associated events

Recall from *Chapter 10* that Lisman (1999) called our attention to the fact that many events that we need to associate in LTM are separated in time by much longer than the ~100 ms activation time course of the CA3 NMDA receptor. Thus, whereas NMDA receptor dynamics could be the basis for associating location A with B and B with C as a rat runs along a path, many more than 100 ms would have elapsed, for example, between when the hapless Jerry dropped his candy and when the opportunistic monkey snatched it up. How might these two events be associated in LTM? It could be that the sustained hippocampal and prefrontal activity documented by Ranganath, Cohen, and Brozinsky (2005) serves this buffering function (*Tip 11.4*).

Tip 11.4

Hebb (1949) anticipated the need for such a temporary buffer, positing a “dual trace mechanism” by which transiently elevated, recurrent activity between the to-be-associated neural elements would persist until the second, structural trace (i.e., synaptic strengthening) could be established.

Functional connectivity-based analysis of the subsequent memory effect

The final reason for devoting so much space to the Ranganath, Cohen, and Brozinsky (2005) study is a separate analysis that was applied to this data set, and published as Ranganath, Heller, Cohen, Brozinsky, and Rissman (2005): a functional connectivity-based analysis of the subsequent memory effect. All of the neuroimaging studies that we have considered up to this point in this chapter used the mass-univariate approach, effectively identifying voxels in which greater signal intensity related to stronger memory encoding. Thus, none of them directly assayed the element that is fundamental to all contemporary models of LTM – the standard model, the multiple trace model, reconsolidation models, etc. – which is the *interaction* between MTL regions and other regions of cortex, in the encoding, consolidation, and retrieval of LTM. *Methodology Box 11.1* illustrates, in the context of the Ranganath, Cohen, and Brozinsky (2005) data, one way in which these interactions can be explicitly taken into account.

The years since the mid-2000s have witnessed an explosion of studies of subsequent memory effects and of MTL-neocortical interactions during consolidation. What we've covered here provides a solid foundation from which to tackle this literature, and next we'll move on to consider memory retrieval. On the way, however, we'll take a detour to the streets of London.

THE HIPPOCAMPUS IN SPATIAL MEMORY EXPERTS

A fitting interregnum between the halves of this chapter focused on encoding-into or retrieval-from LTM are the studies, by Eleanor Maguire and colleagues, that demonstrate how both of these processes can influence the macrostructure of the hippocampus. One account in the popular press sets the stage as follows:

Manhattan's midtown streets are arranged in a user-friendly grid. In Paris 20 administrative districts, or *arrondissements*, form a clockwise spiral around the Seine. But London? A map of its streets looks more like

METHODOLOGY BOX

11.1 Measuring functional connectivity with fMRI

The very basis of LTM is assumed to be the association of neural representations, many of them supported in anatomically distal regions. Nonetheless, the massively parallel univariate analyses used in each of the studies reviewed in the first half of this chapter could neither prove nor disprove this assumption. One way to demonstrate these associations is to measure **functional connectivity**, a term used to describe two or more regions whose activity varies in concert, often in a task-dependent manner. Ranganath et al. (2005) provides an example.

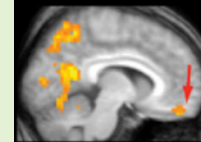
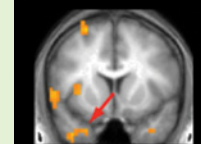
Note in *Figure 11.2.A* that the authors used a “slow” event-related design (see *Tip 11.3*) that partitioned variance in the fMRI signal to four epochs in the trial (probe covariate not shown in the figure). For the functional connectivity analysis, the loadings on each of these trial epoch-related covariates (a.k.a. the parameter estimates, or the beta weights) were used as the data to determine which areas were functionally coupled across trials. Recall that the basis for the subsequent-memory effect is the fact that trial-to-trial fluctuations in a region’s activity level influence subsequent behavior. A functional connectivity analysis evaluates the extent to which those fluctuations are correlated across the brain. Thus, for example, even though the magnitude of activity in areas X, Y, and Z may each predict subsequent memory performance in an experiment, it could be possible that trial-to-trial fluctuations in signal intensity in X are correlated with trial-to-trial fluctuations in Y, but not in Z. The interpretation would be that it is the connectivity between X and Y, in addition to their signal intensity, that underlies successful memory encoding. To the extent that Z plays a role, this must be independent of the other two. Further, it is possible that two other regions, A and B, do not show a univariate subsequent memory effect (i.e., their mean activity level is no higher on subsequently remembered than on subsequently forgotten trials), but that their functional connectivity does. That is, that trials on which the activity in A and B is more tightly coupled are those that produce stronger encoding into LTM. This last point highlights the fact that functional connectivity analyses are multivariate analyses, in that they assay *patterns* of activity in *multiple* voxels.

Back to Ranganath et al. (2005). The authors were specifically interested in whether there were areas for which functional connectivity with the hippocampus predicted stronger encoding into LTM. To assess this, they implemented a “seeded analysis” by selecting the region of left hippocampus identified in *Figure 11.2.A*, extracting from it the early-delay parameter estimates from subsequently remembered trials, and correlating these against the early-delay parameter estimates from all other voxels in the brain. Then they did the same for subsequently forgotten trials. Finally, they determined the areas in which the correlation with the hippocampus was significantly greater on remembered than on forgotten trials. Their findings (*Figure 11.3*) provided direct evidence for something that most assumed, but that few, if any, had demonstrated directly: successful encoding was associated with enhanced functional connectivity between the hippocampus and many neocortical regions within and outside the MTL. Thus, they demonstrated that successful memory formation is associated with transient increases in cortico-hippocampal interaction.

A Cue period



B Early delay



C Late delay

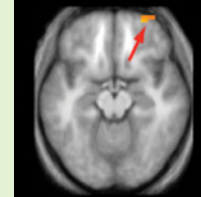


FIGURE 11.3 Seeded functional connectivity analysis, illustrating regions, by trial epoch, whose activity was more strongly correlated with the left hippocampus on trials presenting subsequently remembered vs. subsequently forgotten sample stimuli. Source: Ranganath, Heller, Cohen, Brozinsky, and Rissman, 2005. Reproduced with permission of Wiley.

a tangle of yarn that a preschooler glued to construction paper than a metropolis designed with architectural foresight. Yet London's taxi drivers navigate the smoggy snarl with ease, instantaneously calculating the swiftest route between any two points. (Jabr, 2011)

What does this have to do with the hippocampus? Maguire had noted, in addition to what we already know about experience-dependent synaptic change in the hippocampus, that the hippocampi of some species of food-caching birds undergo seasonal fluctuation in volume, expanding when demands on spatial abilities are greatest. She hypothesized, therefore, that a similar phenomenon might occur in humans whose job requires the memorization, and daily interrogation, of an incredibly detailed mental map of the streets of the city of London. Although several tests of this idea have now been published, *Research Spotlight 11.2* and accompanying *Figure 11.4* focus on perhaps the most tightly controlled of these, a comparison of the brains of London taxi drivers and London bus drivers. The two groups are well matched in terms of socioeconomic status, number of hours per day spent seated while pushing foot pedals and turning a wheel by hand, number of daily interactions with strangers, etc. Importantly, they are also well matched for specific perceptual and physiological factors known to affect the hippocampus, including self-induced movement (linked to the hippocampal theta rhythm), optic flow, and stress. Two salient differences, however, are that (1) taxi cab drivers have undergone intensive training to acquire “The Knowledge” – the layout of 25,000 streets in the city and the locations of thousands of places of interest – and (2) the daily activity of taxi drivers entails planning and navigating in reference to this 2-D mental map, whereas bus drivers follow 1-D routes, often the same one(s) over and over. (Another kind of “memory expert” is profiled in *Web Link 11.1*.)

RETRIEVAL

In cognitive psychology, an influential framework for organizing different kinds of memory retrieval has been Tulving's (1985) distinction between *autonoetic*, *noetic*, and *anoetic* states of awareness with respect to memory retrieval. Derived from a Greek word appropriated by philosophers to refer to “mind” or “intellect,” the term “noetic” in this context roughly

corresponds to “knowing.” Thus autonoetic (or “self-knowing”) awareness refers to an instance of memory retrieval that “is not only an objective account of what has happened or what has been seen or heard ... [but also] necessarily involves the feeling that the present recollection is a reexperience of something that has happened before” (Wheeler, 2000, p. 597). (Note that the concept of “recollection” will take on a technical definition as the parameter in a model of recognition memory that we will consider in some detail further along in this chapter.) Noetic (“knowing”) awareness can occur in two contexts. One is when retrieving information from semantic memory, for which there is not an associated “something that has happened before,” the other is having a “feeling of familiarity” when encountering something or someone, but not being able to retrieve the episode in which this thing or person may have previously been encountered. (“Familiarity,” too, will come up in the context of formal models of recognition memory.) Finally the concept of anoetic (“not knowing”) retrieval captures the fact that nondeclarative memory can be expressed without the individual's awareness that his or her performance is being influenced by a prior experience. In this section, we will consider each of these types of memory retrieval, beginning with how a misattribution of memory-test performance to noetic vs. anoetic memory retrieval can lead to erroneous conclusions about the bases of the amnesic syndrome.

Retrieval without awareness

In the modern era, the initial demonstrations of spared learning in amnesia were generated by Milner and her students and colleagues. First, in 1962, came the report of H. M.'s performance on a test of mirror tracing. This task required the subject to trace out the shape of a star on a piece of paper, while seated at an apparatus that required the paper to be viewed via its reflection off a mirror, because a barrier occluded its direct observation. Although it sounds simple, it's maddeningly difficult at first, as one repeatedly watches one's own hand pushing the pencil in directions that seem to bear no relation to one's volition. Over time, however, one learns the novel sensorimotor mapping imposed by this apparatus (a kind of sensorimotor adaptation largely mediated by the cerebellum), and the number

RESEARCH SPOTLIGHT

11.2 Regionally specific experience-dependent changes in the hippocampus

Bus and taxi drivers were compared with a method called voxel-based morphometry (VBM) that produces measures that can estimate the density of gray matter or the volume of a structure. VBM is typically calculated from T1-weighted anatomical MR images. Importantly, VBM calculations are made after differences in the shape of the structure in question are removed, by calculation of a deformation field (Ashburner and Friston, 2000). Thus, what VBM compares is T1-weighted signal intensity values in each voxel of gray matter – effectively, the average “darkness” of each voxel in group A vs. group B. Differences in VBM are often summarized as differences in “volume.” For example, Maguire, Woollett, and Spiers (2006) summarize the findings of Maguire et al. (2000) as follows:

Examination of [MRI] scans using whole-brain [VBM] ... showed greater gray matter volume in the posterior hippocampi, and reduced anterior hippocampal gray matter volume in taxi drivers when compared with an age-matched control group. In addition, the longer taxi drivers navigated in London the greater the posterior hippocampal gray matter volume, and the more decreased the anterior gray matter volume. (p. 1091)

It should be noted, however, that there is some controversy about exactly what it is that VBM measures, and the extent to which it might be sensitive to “microstructural” differences in the anatomy of different subjects. The Web site of a suite of neuroimaging software tools called FSL includes this caveat on its VBM page:

It is sometimes not possible to determine if the results you may find are the consequence of an effective reduced thickness or atrophy in the gray matter, or rather an indirect reflection of a different gyrification pattern. Indeed, it might be possible that a misalignment of the gyri/sulci or even different folding patterns may lead to the difference of gray matter distribution that you have found.

(Maguire et al. (2000) corroborated their VBM finding with a separate “pixel counting” analysis in which an observer literally measured the surface area of hippocampus in each slice, and volumes were computed from these measurements.) (*Tip 11.5.*)

Note from *Figure 11.4* that taxi drivers had greater gray matter volume than bus drivers in mid-posterior hippocampus, and the reverse in anterior hippocampus. (Nadel, Hoescheidt, and Ryan (2013) drew on this when emphasizing the rostrocaudal gradient in hippocampal functioning that was considered in *Chapter 10.*) The latter can be interpreted in terms of the caudal-to-rostral gradient of dorsal-stream input into the hippocampus, in that it would be expected that parietal and parahippocampal inputs drive the formation of highly detailed spatial maps in the hippocampus. The anterior hippocampus effects are not as easy to interpret. They are the result of smaller VBM values in the taxi group, not larger values in the bus group (e.g., the same pattern had been observed in the 2000 study, in which the controls were not professional drivers). Paralleling this neuroanatomical finding was the fact that the taxi group was significantly impaired on a test of delayed recall for an abstract, complex figure, a task that is sensitive to right anterior hippocampal damage. Thus, one possible interpretation is that the unique circumstances of being a London taxi driver produce a trade-off of regional hippocampal processing power.

Research Spotlight 11.2 *Continued*

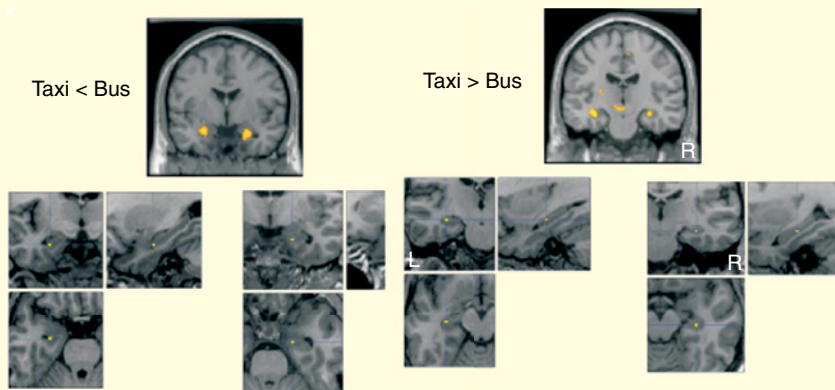


FIGURE 11.4 Differences in gray matter volume in the hippocampi of taxi cab drivers vs. bus drivers. The top image in both the Taxi > Bus and Taxi < Bus panels displays the effects on a full coronal slice from an atlas, and the images below show, in coronal, sagittal, and axial views, on the higher-resolution structural image of a single subject from the study, the location of the voxel from each hemisphere showing the largest VBM difference at the group level. Source: Maguire, Woollett, and Spiers, 2006. Reproduced with permission of Wiley.

Tip 11.5

It will be interesting to see if postmortem exams (which I'm assuming Dr. Maguire has arranged for) reveal that the taxi drivers also have, on average, a greater density of neurons and/or of dendritic branching in the posterior hippocampus.

of errors declines with each trial. Impressively, H. M. also showed a steady decline in errors across trials, and, astoundingly, his first trial on day 2 of testing featured many fewer errors than he had made at the beginning of day 1! By day 3, his performance on the task was near error-free, this despite the complete absence of any conscious memory of ever having performed the task. Soon thereafter it was confirmed that H. M. demonstrated preserved motor learning on many motor performance-based tasks.

H. M. also demonstrated learning on what came to be called repetition priming tasks. For example, a set of 20 line drawings of common objects was modified to create four incomplete versions of each, with version 1 resembling a series of arbitrary dashes and marks on

the page – very difficult to recognize – and each subsequent version containing progressively more details. Subjects were first shown version 1 of each image, and encouraged to name it as quickly as possible. If they didn't get all 20 correct (and no individual did), they were then shown all the images from version 2, and so on, until they correctly named each drawing from a version. H. M. correctly identified 6 from version 1, 12 from version 2, 19 from version 3, and all 20 from version 4. An hour later, he was retested on the same procedure, and this time correctly identified 14 from version 1, 15 from version 2, and all 20 from version 3, overall making 48% fewer errors. Although this was a marked learning effect, it was characterized by Milner, Corkin, and Teuber (1968) as “vestigial,” and was smaller than that of age- and education-matched control subjects, whose improvement ranged from 52% to 93%. “All in all,” they concluded, “we remain impressed with the central role of the learning impairment in these patients” (p. 233). Around this time, however, another group was interpreting very similar findings in a radically different way. *Thought Box 11.1* outlines how consideration of awareness at the time of retrieval contributed to the resolution of this debate.

THOUGHT BOX

11.1 Encoding- vs. retrieval-based accounts of anterograde amnesia

Logically, the phenomenon of relatively spared retrograde memory after bilateral damage to the MTL does not, alone, permit one to distinguish between a “consolidation block” (i.e., impaired encoding) account versus disordered storage or disordered retrieval accounts of MTL function. This was illustrated most famously by British neuropsychologist Elizabeth Warrington and British neuroscientist Lawrence Weiskrantz, with a series of papers in the late 1960s and early 1970s. A 1974 study, for example, highlighted performance on two memory tests administered 10 minutes after the presentation of 16 words. Amnesic patients performed as well as control subjects on what the authors termed a “cued recall” task, in which the first three letters of a studied word (so-called “word stems”) were presented and the subject was “required to identify the stimulus word” (p. 420). The patients were markedly impaired, however, on a test of Yes/No recognition. These results, together with evidence of disproportionate sensitivity to interference from items presented prior to or after the critical information, led Warrington and Weiskrantz (1974) to argue that MTL amnesia was better characterized by “altered control of information in storage” (p. 419), rather than a consolidation block. That is, they argued that new information “gets in there” (my quotes), but that the amnesic syndrome makes it more difficult to access this information.

Subsequent research has established that the flaw in Warrington and Weiskrantz’s reasoning was their failure to appreciate the distinction between noetic vs. anoetic states of memory retrieval. With word-stem completion, for example, it has subsequently been shown that the performance of amnesic subjects relative to control subjects depends on the precise phrasing of instructions about how to process the three-letter stems: when subjects are instructed to complete the three-letter stems to “the first word that comes to mind,” and no reference is made to the prior study episode, amnesic patients often generate target words at a level that is comparable to that of control subjects. When, in contrast, they are instructed to use the three-letter stem as a cue with which to retrieve an item from the studied list, amnesic patients are typically impaired. The former procedure, which more closely resembles that used by Warrington and Weiskrantz (1974), has come to be known as word-stem completion *priming*, the latter as word-stem *cued recall*. Thus, the intact performance of amnesic patients in the Warrington and Weiskrantz (1974) study has come to be reinterpreted as an early demonstration of intact performance by amnesic subjects on a priming task, a phenomenon that falls under the rubric of *nondeclarative memory*. (For Further Reading, see Squire, 2006.)

The term “priming” is a metaphor referring to priming a pump – filling it with fluid so that when subsequently turned on (or activated via its handle) it will start working more quickly. With the psychological phenomenon of repetition priming, for example, the initial visual presentation of a stimulus makes it easier for the visual system to process that stimulus on subsequent repeat presentations. This comes about as a result of plasticity in the cortical systems that process these stimuli, regardless of whether or not there is an intact MTL to make a record of this processing.

Documenting contextual reinstatement in the brain

“Our results ground Tulving’s speculations about mental time travel in neural fact” (Polyn, Natu, Cohen, and Norman, 2005, p. 1966).

Now *that’s* what I’m talkin’ about! (Tip 11.6.)

The study to which we now turn speaks directly to the neural instantiation of autonoetic retrieval. It was motivated by the idea from cognitive psychology that the British psychologist Sir Frederick Bartlett (1886–1969) had articulated in 1932, and that Tulving had codified in 1973 as “contextual reinstatement”: when asked to recall a particular memory, people first activate knowledge about the general properties of the event in question, and then use this general knowledge to constrain the search for the memory of the target event. To assess the neurobiological plausibility of this “speculation,” Sean Polyn, his then-graduate mentor Kenneth Norman (himself a mentee of Schacter and O’Reilly), and their colleagues, devised an innovative application of multivariate pattern analysis (MVPA). The resultant publication (Polyn et al., 2005) was one among a handful of seminal publications, all published within a year or so of each other, that brought to the attention of the broader cognitive neuroscience community just how powerful this technique could be.

The dynamics of multivariate state changes underlying behavior

The experiment of Polyn et al. (2005) proceeded in three phases. The first was the study phase, during which subjects were scanned while viewing 90 stimuli, 30 drawn from each of three categories: famous people; famous locations; and common

household objects. This was an incidental encoding task in that, for each stimulus, subjects were requested to indicate either (a) how much they liked the famous person; (b) how much they would like to travel to the famous location; or (c) how often they use the common object. (“*But wait,*” you ask, “*I thought this section was about memory retrieval. So why this study in which they scan during encoding?*” “*A perceptive question,*” I respond, “*we’ll see why in a moment.*”)

The second phase of the experiment was to train a neural network pattern classifier to discriminate brain activity from each of the three categories. The process of training the classifier was not different, in principle, from how Rogers, Hodges, Lambon Ralph, and Patterson (2003; *Research Spotlight 9.1*) trained their PDP model to recognize verbal and visual objects: they presented it with the vector of signal intensity values produced by a given stimulus item, evaluated the network’s guess as to which category the stimulus belonged to, and, when appropriate, propagated an error signal through the network that adjusted its weights such that it would be more likely to correctly classify that pattern of signal intensity values in the future. Once the classifier was fully “trained up,” it could be fed the data from phase 3.

Phase 3 was a free recall session, performed in the scanner, in which subjects were simply instructed to recall as many items from the study phase as possible, by speaking them into an MR-compatible microphone. Data from one subject are shown in *Figure 11.5*, and the first thing you may notice is that items were recalled in clusters: first, the subject recalled ~8 locations; then ~6 celebrities (“faces”); then ~6 objects; and so on. This kind of clustering in recall has been known to psychologists for decades and, indeed, is one of the behavioral phenomena consistent with a contextual reinstatement mechanism. What’s really amazing about this figure is the relation that it reveals between this distinctive pattern of free recall performance and the concurrent fluctuations in neural activity: just prior to each “recall cluster,” one can see a reconfiguration of brain activity into the pattern corresponding to that cluster’s category! That is, we see the “neural context” associated with thinking about, let’s say, famous locales, getting *reinstated* in the subject’s brain just before she comes out with “Taj Mahal ... Eiffel Tower ... Machu Pichu.” Thus, an

Tip 11.6

This summative statement from Polyn and colleagues (2005) captures a big part of the appeal of cognitive neuroscience: it allows for the objective assessment of which ideas – often from psychology and philosophy – about how the mind *might* work, or how it *should* work, are likely to be true based on what we can learn about how the brain *does* work.

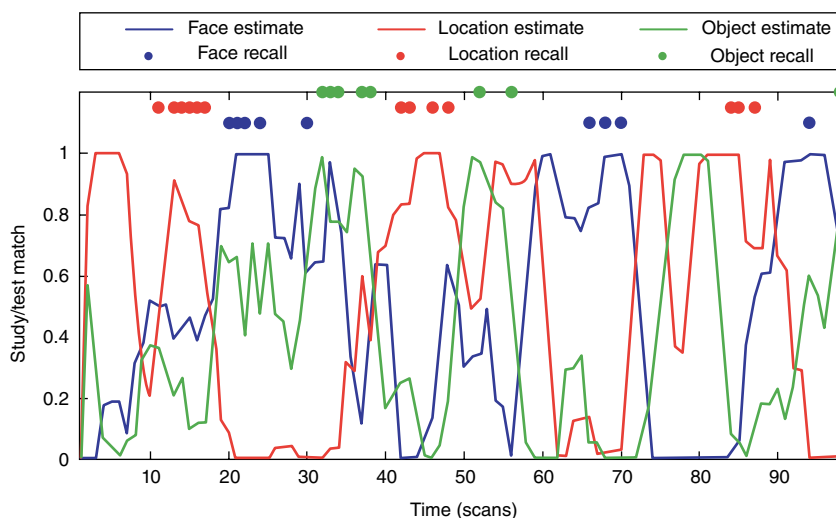


FIGURE 11.5 Contextual reinstatement during free recall. The blue, red, and green waveforms indicate, at each point in time during a scan of a single subject, the extent to which neural activity matches the pattern of neural activity independently determined to correspond to that stimulus category. Each dot indicates the recall of an item (larger dots for the recall of multiple items within the same 1.8 second volume acquisition; dots have been shifted forward in time by three TRs (i.e., 5.4 s) to account for the hemodynamic lag). Note how each instance of recall is preceded by brain activity taking on that category's pattern. Source: Polyn, Natu, Cohen, and Norman, 2005. Reproduced with permission of AAAS.

important component of autonoetic awareness in retrieval is the reconfiguration of brain activity into the state that it was in when encoding the to-be-remembered information. (More on this study at *Web Clip 11.1*.)

Implications for “mind reading” ... well, “brain reading”

Aside from what it tells us about memory retrieval, the Polyn et al. (2005) study garnered considerable excitement because it demonstrated a primitive form of “brain reading.” Particularly intriguing was that, unlike the examples of MVPA that we’ve considered up to this point, in this study the experimenters were not demonstrating that a classifier could correctly guess what was the independent variable on a particular trial (e.g., “on this trial the stimulus was a picture of a cat”). Rather, their classifier was tracking the evolution of a mental *dependent* variable – memory recall – that was unaccompanied by any external cues or special instructions. I qualify it as primitive, because we have to keep in mind that this demonstration occurred within a highly constrained environment in

which the classifier only “knew about” three states in the world – celebrities, famous locales, and common objects. Thus, if Polyn’s subjects ever thought about global warming, or that evening’s social plans, or how uncomfortable they were lying in the scanner, the classifier would only be able to assess whether the corresponding pattern of activity was more celebrity-like, locale-like, or object-like (*Tip 11.7*).

Tip 11.7

Studies like that of Polyn et al. (2005) with MVPA decoding models set the stage for the development of more ambitious techniques, such as multivariate *encoding* models, that build models of how the brain encodes information such that they can, for example, reconstruct from fMRI data what stimulus a subject must have been looking at, even if the subject had never before seen that stimulus. (See *Further Reading* reference to Serences and Saproo, 2012.)

In addition to its technical prowess and verging-on-SciFi evocations, this approach can also be seen as offering possible ways forward for understanding many unresolved questions about neurophysiological and neurochemical mechanisms underlying memory retrieval. One very salient example is reconsolidation (*Real-World Applications 10.1*), a phenomenon that is well-documented, but remains poorly understood. A second is the phenomenon of state dependence in memory retrieval: we'll see in *Chapter 17's* treatment of emotional memory that there is neuropharmacological evidence that one mechanism underlying memory retrieval is that the brain "puts itself" into a physiological state resembling the state that it was in when encoding took place.

Familiarity vs. recollection

Dual-process models

We'll conclude this chapter with a consideration of whether the theoretical distinction between auto-noetic and noetic retrieval corresponds to qualitatively, or just quantitatively, different neural mechanisms. There are numerous, by and large quite similar, theoretical models from cognitive psychology that can be classified as dual-process models. These posit that free recall depends on a process called **recollection**, and that recognition memory can be supported by recollection and/or by a process called **familiarity**. According to these theories, recollection entails the retrieval of qualitative information associated with the learning episode. For example, when I recollect what I ate for dinner last night (it was my twentieth wedding anniversary! really), I remember not only the scallops that I ate, but also where we sat at the restaurant, what the weather was like outside, the dress that my lovely dinner companion was wearing, etc. On a more mundane level, subjects in the Polyn et al. (2005) study, for example, would be assumed to have recollected contextual information about the moments when each of the images that they recalled had been presented at study. (For example, perhaps, when recalling having seen an image of the actor John Wayne, they also recollected how they had rated John Wayne during the study phase.) Recollection is characterized as a relatively slow process.

Familiarity, in contrast, is a fast, quantitative memory signal that only conveys the information that one has or has not previously encountered the item in question. Let's return again to the author's personal life, and engage in some mental time travel. Let's say that later today, when I need to take a break from writing, I will go jogging in my neighborhood and will encounter, walking along the sidewalk, a woman and, few minutes later, a man. As I run past the woman I have a strong sense that I've seen her somewhere recently, but where was it? Being the omniscient narrator as well as the protagonist in this imagined anecdote, I can tell you that this woman was the waitress from last night's dinner, but that I didn't recognize her in this very different context. Subsequently, as I run past the man, I have a vague sense, but less strong, that he, too, "looks familiar." He, it turns out, drove the taxi cab that took us to the restaurant. (Or was he the driver from the ride home? ...) Thus, familiarity can vary along the dimension of strength, whereas recollection is assumed to be an all-or-none phenomenon. (Recollected memories can, of course, contain more or less detail, but this has more to do with the stored memory than the process of retrieval, per se.)

And these two processes are assumed to work in tandem. As it turns out, I actually did go out for a run this afternoon, and although I encountered neither the waitress nor the cabbie, I did run past someone worthy of note. He was coming the other direction along the Lakeshore Path, and at first looked like just another afternoon jogger. Then, I thought, as we were passing, "that guy looks familiar." After another three or four strides it came back to me: *His name is Charlie, he's in the Business School (Marketing Department?), and he and I used to serve on a committee that met 10 or so times each Fall semester.* By the dual-processes account, the visual processing of Charlie coming toward me along the path first generated a familiarity signal, which, after a few seconds of cogitation, gave way to a recollection.

As stated previously, there are many dual-process models. There have also been many, many behavioral, ERP, and fMRI studies testing different facets of these models. A general pattern that has emerged from this literature is that recollection depends on the hippocampus proper, whereas familiarity is supported by neocortical regions of the MTL, particularly

the perirhinal cortex. We'll conclude this chapter by illustrating just one narrow slice from this literature, highlighting the methods underlying one particular model – the **receiver operating characteristic (ROC)** –based model – and a study that addresses the role of the hippocampus in these two hypothesized processes.

Dual processes of memory retrieval in the rodent

One of the simplest forms of recognition test is “yes/no” recognition, in which subjects indicate whether or not they have seen the memory probe previously. In two-alternative forced choice, an old and new item are presented together, and the subject must indicate which is old (or, in the case of delayed nonmatch-to-sample (DNMS), which is new). Although both of these procedures assess memory performance, neither addresses what underlying processes may contribute to this performance. Several procedures for estimating the relative contribution of familiarity- vs. recollection-based processes to recognition performance do so empirically. For example, in the “remember/know” procedure, instead of responding “yes” or “no” to the probe, subjects indicate whether they “remember” having seen it, in which case they are assumed to have used recollection, or whether they only “know” that they have seen it, but can't retrieve a memory of having done so, in which case they are assumed to have used familiarity. The Wagner et al. (1998) and Brewer et al. (1998) studies with which we opened this chapter each effectively used this procedure, by asking subjects to rate their confidence for each positive response, and only using high-confidence hits in their subsequent-memory analyses.

An alternative procedure is to derive the contribution of the two processes to recognition performance by estimating ROC curves from one's subjects. This is the approach taken by Fortin, Wright, and Eichenbaum (2004) in their study of recognition memory in the rat, and its fundamentals are introduced *Methodology Box 11.2*.

How did Fortin and colleagues manage to get rats to give a range of five levels of confidence judgments? Let's start with their basic procedure. Rats were trained to retrieve rewards (pieces of the breakfast cereal Cheerios) buried in cups of sand that was

scented with one of 30 distinctive household scents. Next, they learned the DNMS procedure by which, if a test scent was a new scent, they were to dig in that sand to retrieve the reward. If, on the other hand, it was an old scent, the cup would not contain a reward, but the rat would be rewarded for refraining from digging and, instead, moving to the back of the cage to an alternate, empty cup to which a reward would be delivered. Different levels of bias were then created by, in effect, varying how hard the rat would have to work to receive its reward for correctly choosing “new” vs. how much food it would receive for correctly choosing “old.” A liberal bias was created by making it more onerous to respond “new,” by placing the same amount of sand in a taller cup and reducing the amount of reward that it would contain. Thus, if the animal “wasn't sure” about whether a new scent was old or new, it would tend to favor an “old” response, because a “new” response required too much effort. (If, however, it was certain that the test scent was new, then the only choice was to climb and dig.) In tandem, the amount of reward delivered for correct “old” responses was increased, thereby further incentivizing this choice.

Now let's look at the results, in the bottom three panels of *Figure 11.6*. Panel D shows the preoperative performance of the animals that were to receive hippocampal lesions, and of the control animals. The combined ROC of the two groups shows the standard asymmetric shape, and the model estimates (inset bar graphs) indicate comparably strong familiarity and recollection estimates for both groups. Panel E illustrates the key result: after hippocampal lesions, the estimated contribution to performance of recollection is devastated, but that of familiarity is intact. The dotted line labeled “Con F” shows the estimated performance of the control group when the contribution of recollection is algebraically removed. Finally, panel F shows the effect, in the control rats, of lengthening the delay period from 30 to 75 minutes. This manipulation removes the contribution of familiarity to their performance but leaves recollection relatively intact. Not shown in this figure is the effect of the hippocampal lesion on overall performance: the control animals at the 30 minute delay performed at 73% correct, the hippocampal animals at 66% correct (i.e., a modest impairment) and the

METHODOLOGY BOX

11.2 Estimating memory signals with an ROC-based procedure

The receiver operating characteristic (ROC) gets its arcane name from the fact that it was first developed during World War II to understand the performance of radar operators whose difficult (and consequential) task was to observe a noisy screen on the radar receiver and to detect signals (e.g., blips corresponding to enemy aircraft) from among the noise. They faced the classic tradeoff that was later formalized in **signal detection theory**: by setting too lenient a criterion the operator risked issuing “false alarms” – announcing that he had detected enemy aircraft when, in fact, it was just noise – whereas, by setting too strict a criterion, he risked missing actual targets. Note that, in this scenario, the radar operator can change his performance by “setting his criterion” to be liberal or to be strict, but none of this changes the inherent signal-to-noise ratio (SNR) of the radar receiver. Further, a highly trained operator would better discriminate targets from noise than would a novice.

Applied to recognition memory, all items that one encounters are assumed to generate some recognition signal, and one’s “mnemonic SNR” is the extent to which *old* items generate a stronger signal than *new* items (*Tip 11.8*). Independent of one’s SNR is where one chooses to set one’s criterion. The tradeoff associated with different criterion settings is implemented in an ROC-based recognition test by requiring subjects to rate each recognition probe on a scale from 1 to, say, 5, with “5” corresponding to “I’m absolutely certain that I saw this previously” and “1” corresponding to “I’m certain that I did not see this previously.” After collecting the data, one can then derive an ROC curve in the following way. First, set up a pair of axes like those illustrated in *Figure 11.6*, with the probability of hits along the vertical axis and the probability of false alarms along the horizontal axis. Next, score the subject’s data as though only his “5” responses were “yes”es and all his other responses were “no”s. This estimates his performance at its most conservative: his hit rate will be the lowest, but so, too, will his false-alarm rate. Next, rescore the data as though responses of “5” and “4” were “yes”es, and the others “no”s. Because he had assigned 4s to probes about which he had relatively high confidence, the “5s + 4s” hit rate will be higher than was the “5s only” hit rate. But so, too, will be the false alarm rate. Thus, the “5s + 4s” data point will be higher along both the hits axis and the false alarms axis. The progression from most conservative to most liberal can be seen in panel D of *Figure 11.6*.

Panel A of *Figure 11.6* is an idealized plot of ROC-analyzed data from a human subject in a word-list recognition experiment. When this “asymmetric ROC” is decomposed, as in panels B and C, it can be understood as being made up of independent factors represented by each of these two: the degree of curvilinearity of the function labeled “familiarity,” and the y-intercept of the function labeled “recollection” (Yonelinas, 2001). Note that overall memory performance is captured by the area under the curve, with the diagonal equivalent to chance performance.

Tip 11.8

What’s informally referred to here as “mnemonic SNR” is formalized in signal detection theory as the measure d' (pronounced *dee prime*).

control animals at the 75 minute delay performed at 64% correct. (Note that whereas lesioning the hippocampus removes the estimated contribution of recollection to performance, increasing the delay by and large removes the estimated contribution of familiarity to performance.)

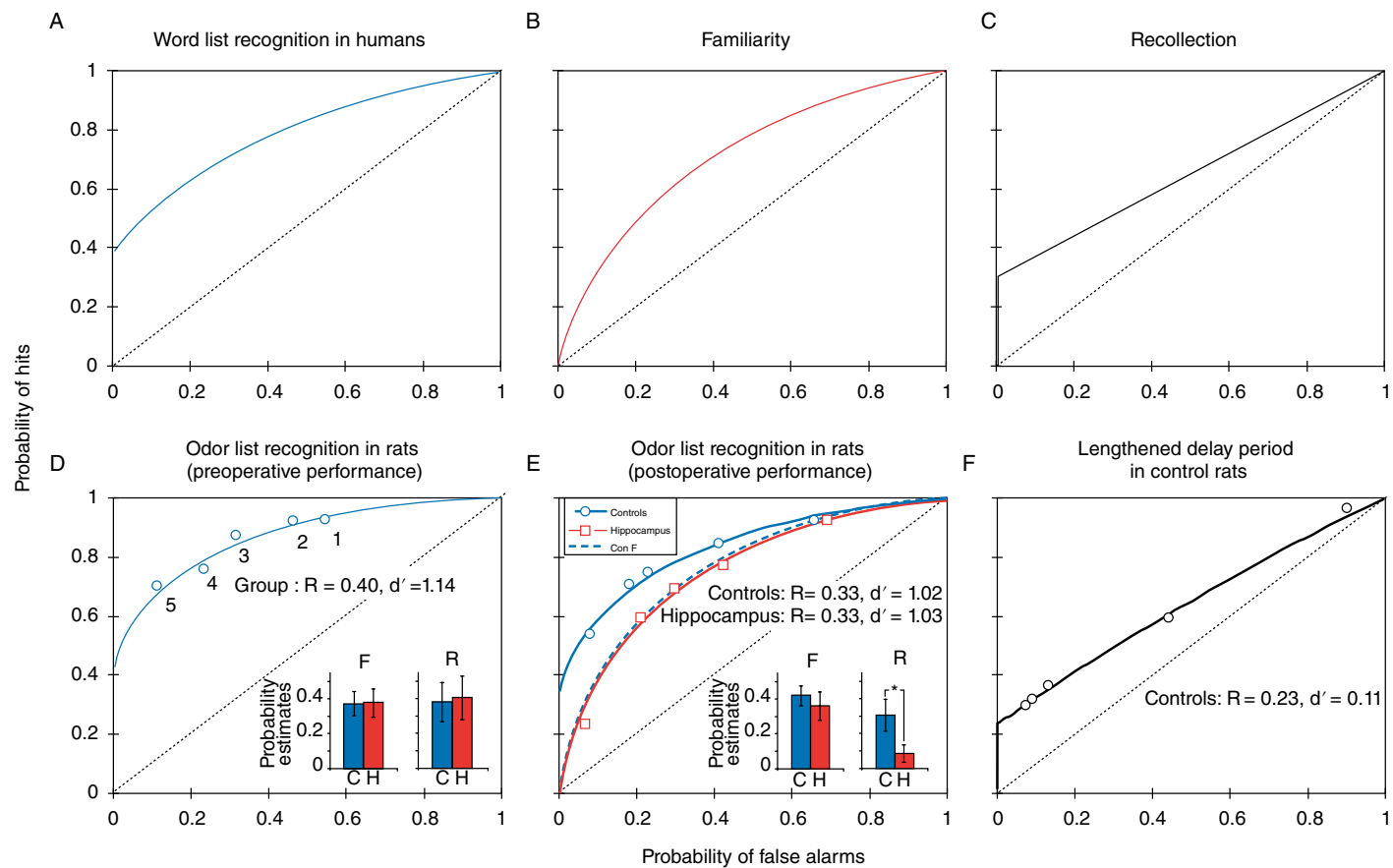


FIGURE 11.6 ROC-based dual-process model of recognition, applied to humans and rats. Source: Fortin, Wright, and Eichenbaum, 2004. Reproduced with permission of Nature Publishing Group.

This suggests a possible solution to why the monkeys with H lesions described in the Murray, Bussey, and Saksida (2007) review (*Figure 10.7*) were not impaired on tests of DNMS: perhaps they could solve the tasks based on a familiarity signal.

KNOWLEDGE

That's the topic of our next chapter. Having considered the principles of how information gets into, and gets retrieved from, the brain, we next turn to how it is organized "while it sits there," waiting to be accessed and thought about.

END-OF-CHAPTER QUESTIONS

1. Why can one not "just trial-average" stimulus-evoked fMRI signal as a function of subsequent memory performance, as can be done with ERP data?
2. Is it possible to learn anything about LTM encoding from a neuroimaging study of STM?
3. What principle of LTM formation is better captured with functional connectivity analyses than with signal intensity-based analyses?
4. What is it about the experience/expertise of London taxi drivers, and the organization of cortical projections to the hippocampus, that accounts for the pattern of structural change observed in these individuals?
5. Define auto-noetic, noetic, and anoetic memory retrieval, and describe a task that would be well suited to assess each.
6. Would it be possible to document contextual reinstatement in the brain with univariate analyses of fMRI data? Explain how, or why not.
7. Is there neural evidence for a distinction between familiarity and recollection in memory retrieval? To what kind of task is it assumed that only one of these can contribute, and to what can both contribute?

REFERENCES

- Ashburner, John, and Karl J. Friston. 2000. "Voxel-Based Morphometry – The Methods." *NeuroImage* 11 (6): 805–821. doi: 10.1006/nimg.2000.0582.
- Bartlett, Frederic C. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Brewer, James B., Zuo Zhao, John E. Desmond, Gary H. Glover, and John D. E. Gabrieli. 1998. "Making Memories: Brain Activity that Predicts How Well Visual Experience Will Be Remembered." *Science* 281 (5380): 1185–1187. doi: 10.1126/science.281.5380.1185.
- Dale, Anders M., and Randy L. Buckner. 1997. "Selective Averaging of Rapidly Presented Individual Trials Using fMRI." *Human Brain Mapping* 5 (5): 329–340. doi: [http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1997\)5:5<329::AID-HBM1>3.3.CO;2-A](http://dx.doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.3.CO;2-A).
- Fortin, Norbert J., Sean P. Wright, and Howard Eichenbaum. 2004. "Recollection-Like Memory Retrieval in Rats Is Dependent on the Hippocampus." *Nature* 431 (7005): 188–191. doi: 10.1038/nature02853.
- Hebb, Donald Olding. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley and Sons.
- Jabr, Ferris. 2011. "Cache Cab: Taxi Drivers' Brains Grow to Navigate London's Streets." *Scientific American*, December 8. <http://www.scientificamerican.com/article.cfm?id=london-taxi-memory>.
- Kanwisher, Nancy, Josh McDermott, and Marvin M. Chun. 1997. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception." *Journal of Neuroscience* 17 (11): 4302–4311. <http://www.jneurosci.org/content/17/11/4302.full.pdf+html>.
- Lisman, John E. 1999. "Relating Hippocampal Circuitry to Function: Recall of Memory Sequences by Reciprocal Dentate–CA3 Interactions." *Neuron* 22: 233–242. doi: 10.1016/S0896-6273(00)81085-5.

- Maguire, Eleanor A., David G. Gadian, Ingrid S. Johnsrude, Catriona D. Good, John Ashburner, Richard S. J. Frackowiak, and Christopher D. Frith. 2000. "Navigation-Related Structural Change in the Hippocampi of Taxi Drivers." *Proceedings of the National Academy of Sciences of the United States of America* 97 (8): 4398–4403. doi: 10.1073/pnas.070039597.
- Maguire, Eleanor A., Katherine Woollett, and Hugo J. Spiers. 2006. "London Taxi Drivers and Bus Drivers: A Structural MRI and Neuropsychological Analysis." *Hippocampus* 16 (12): 1091–1101. doi: 10.1002/hipo.20233.
- Milner, Brenda, Suzanne Corkin, and H.-L. Teuber. 1968. "Further Analysis of the Hippocampal Amnesic Syndrome: 14-Year Follow-Up Study of H. M." *Neuropsychologia* 6 (3): 215–234. doi: 10.1016/0028-3932(68)90021-3.
- Morris, Richard G. M., E. Anderson, Gary S. Lynch, and Michael Baudry. 1986. "Selective Impairment of Learning and Blockade of Long-Term Potentiation by an N-Methyl-D-Aspartate Receptor Antagonist, AP5." *Nature* 319 (6056): 774–776. doi: 10.1038/319774a0.
- Murray, Elisabeth A., Timothy J. Bussey, and Lisa M. Saksida. 2007. "Visual Perception and Memory: A New View of Medial Temporal Lobe Function in Primates and Rodents." *Annual Review of Neuroscience* 30: 99–122. doi: 10.1146/annurev.neuro.29.051605.113046.
- Nadel, Lynn, Siobhan Hoscheidt, and Lee R. Ryan. 2013. "Spatial Cognition and the Hippocampus: The Anterior-Posterior Axis." *Journal of Cognitive Neuroscience* 25 (1): 22–28. doi: 10.1162/jocn_a_00313.
- Petersen, Steven E., Peter T. Fox, Michael I. Posner, Mark Mintun, and Marcus E. Raichle. 1988. "Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing." *Nature* 331 (6157): 585–589. doi: 10.1038/331585a0.
- Polyn, Sean M., Vaidehi S. Natu, Jonathan D. Cohen, and Kenneth A. Norman. 2005. "Category-Specific Cortical Activity Precedes Retrieval During Memory Search." *Science* 310 (5756): 1963–1966. doi: 10.1126/science.1117645.
- Ranganath, Charan, Michael X. Cohen, and Craig J. Brozinsky. 2005. "Working Memory Maintenance Contributes to Long-Term Memory Formation: Neural and Behavioral Evidence." *Journal of Cognitive Neuroscience* 17 (7): 994–1010. doi: 10.1162/0898929054475118.
- Ranganath, Charan, Aaron Heller, Michael X. Cohen, Craig J. Brozinsky, and Jesse Rissman. 2005. "Functional Connectivity with the Hippocampus During Successful Memory Formation." *Hippocampus* 15 (8): 997–1005. doi: 10.1002/hipo.20141.
- Rogers, Timothy T., John R. Hodges, Matthew A. Lambon Ralph, and Karalyn Patterson. 2003. "Object Recognition Under Semantic Impairment: The Effects of Conceptual Regularities on Perceptual Decisions." *Language and Cognitive Processes* 18 (5–6): 625–662. doi: 10.1080/01690960344000053.
- Stern, Chantal E., Suzanne Corkin, R. Gilberto González, Alexander R. Guimaraes, John R. Baker, Peggy J. Jennings, ... Bruce R. Rosen. 1996. "The Hippocampal Formation Participates in Novel Picture Encoding: Evidence from Functional Magnetic Resonance Imaging." *Proceedings of the National Academy of Sciences of the United States of America* 93 (16): 8660–8665.
- Tulving, Endel. 1985. "Memory and Consciousness." *Canadian Psychology* 26 (1): 1–12. doi: 10.1037/h0080017.
- Tulving, Endel, Hans J. Markowitsch, Shitij Kapur, Reza Habib, and Sylvain Houle. 1994. "Novelty Encoding Networks in the Human Brain: Positron Emission Tomography Data." *NeuroReport* 5 (18): 2525–2528. doi: 10.1097/00001756-199412000-00030.
- Wagner, Anthony D., Daniel L. Schacter, Michael Rotte, Wilma Koutstaal, Anat Maril, Anders M. Dale, Bruce R. Rosen, and Randy L. Buckner. 1998. "Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity." *Science* 281 (5380): 1188–1191. doi: 10.1126/science.281.5380.1188.
- Warrington, Elizabeth K., and Lawrence Weiskrantz. 1974. "The Effect of Prior Learning on Subsequent Retention in Amnesic Patients." *Neuropsychologia* 12 (4): 419–428. doi: 10.1016/0028-3932(74)90072-4.
- Wheeler, Mark A. 2000. "Episodic Memory and Autonoetic Awareness." In *The Oxford Handbook of Memory*, edited by Endel Tulving and Fergus I. M. Craik, 597–608. New York: Oxford University Press.
- Yonelinas, Andrew P. 2001. "Components of Episodic Memory: The Contribution of Recollection and Familiarity." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 356 (1413): 1363–1374. doi: 10.1098/rstb.2001.0939.

OTHER SOURCES USED

- Gabrieli, John D. E., Margaret M. Keane, Ben Z. Stanger, Margaret M. Kjelgaard, Suzanne Corkin, and John H. Growdon. 1994. "Dissociations Among Structural-Perceptual, Lexical-Semantic, and Event-Fact Memory Systems in Amnesia, Alzheimer's Disease, and Normal Subjects." *Cortex* 30 (1): 75–103. doi: 10.1016/S0010-9452(13)80325-5.
- Holyoak, Keith J. 1999. "Psychology." In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil, xxxix–xlix. Cambridge, MA: MIT Press.
- Lee, Diane W., Troy G. Smith, Anthony D. Tramontin, Kiran K. Soma, Eliot A. Brenowitz, and Nicola S. Clayton. 2001. "Hippocampal Volume Does Not Change Seasonally in a Non Food-Storing Songbird." *NeuroReport* 12 (9): 1925–1928. doi: 10.1097/00001756-200107030-00031.
- Moscovitch, Morris. 2000. "Theories of Memory and Consciousness." In *The Oxford Handbook of Memory*, edited by Endel Tulving and Fergus I. M. Craik, 609–625. New York: Oxford University Press.
- Rissman, Jesse, Adam Gazzaley, and Mark D'Esposito. 2004. "Measuring Functional Connectivity During Distinct Stages of a Cognitive Task." *NeuroImage* 23 (2): 752–763. doi: 10.1016/j.neuroimage.2004.06.035.
- Rugg, Michael D., and Kevin Allan. 2000. "Event-Related Potential Studies of Memory." In *The Oxford Handbook of Memory*, edited by Endel Tulving and Fergus I. M. Craik, 521–538. New York: Oxford University Press.
- Smulders, Tom V., A. D. Sasson, and Timothy J. DeVoogd. 1995. "Seasonal Variation in Hippocampal Volume in a Food-Storing Bird, the Black-Capped Chickadee." *Journal of Neurobiology* 27 (1): 15–25. doi: 10.1002/neu.480270103.
- Warrington, Elizabeth K., and Lawrence Weiskrantz. 1968. "New Method of Testing Long-Term Retention with Special Reference to Amnesic Patients." *Nature* 217 (5132): 972–974. doi: 10.1038/217972a0.
- Warrington, Elizabeth K., and Lawrence Weiskrantz. 1970. "The Amnesic Syndrome: Consolidation or Retrieval?" *Nature* 228 (5272): 628–630. doi: 10.1038/228628a0.
- Williams, Pepper, and Michael J. Tarr. 1997. "Structural Processing and Implicit Memory for Possible and Impossible Figures." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23 (6): 1344–1361. doi: 10.1037/0278-7393.23.6.1344.
- Williams, Pepper, and Michael J. Tarr. 1999. "Orientation-Specific Possibility Priming for Novel Three-Dimensional Objects." *Perception and Psychophysics* 61 (5): 963–976. doi: 10.3758/BF03206910.
- Zola, Stuart M., Larry R. Squire, Edmond Teng, Lisa Stefanacci, Elizabeth A. Buffalo, and Robert E. Clark. 2000. "Impaired Recognition Memory in Monkeys after Damage Limited to the Hippocampal Region." *Journal of Neuroscience* 20 (1): 451–463.

FURTHER READING

- Ashburner, John, and Karl J. Friston. 2001. "Why Voxel-Based Morphometry Should Be Used." *NeuroImage* 14 (6): 1238–1243. doi: 10.1006/nimg.2001.0961.
A response to the critique from Bookstein (2001).
- Bookstein, Fred L. 2001. "'Voxel-Based Morphometry' Should Not Be Used with Imperfectly Registered Images." *NeuroImage* 14 (6): 1454–1462. doi: 10.1006/nimg.2001.0770.
One critique of VBM, suggesting circumstances in which face-value interpretation of its results may lead to flawed inference.
- Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby. 2006. "Beyond Mind-Reading: Multi-Voxel Pattern Analysis of fMRI Data." *Trends in Cognitive Sciences* 10 (9): 424–430. doi: 10.1016/j.tics.2006.07.005.
An accessible tutorial review on applications of MVPA decoding models in neuroimaging research.
- Paller, Ken A., and Anthony D. Wagner. 2002. "Observing the Transformation of Experience into Memory." *Trends in Cognitive Sciences* 6 (2): 93–102. doi: 10.1016/S1364-6613(00)01845-3.
Review of the theoretical and ERP literatures on the subsequent memory effect that laid the groundwork for the fMRI studies detailed in this chapter.

- Rissman, Jesse, and Anthony D. Wagner. 2012. "Distributed Representations in Memory: Insights from Functional Brain Imaging." *Annual Review of Psychology* 63: 101–128. doi: 10.1146/annurev-psych-120710-100344.
Review of MVPA analyses of declarative memory.
- Rugg, Michael D., and Andrew P. Yonelinas. 2003. "Human Recognition Memory: A Cognitive Neuroscience Perspective." *Trends in Cognitive Sciences* 7 (7): 313–319. doi: 10.1016/S1364-6613(03)00131-1.
A good place to start if contemplating a plunge into the massive literature on the cognitive neuroscience of dual-process models of retrieval.
- Serences, John T., and Sameer Saproo. 2012. "Computational Advances towards Linking BOLD and Behavior." *Neuropsychologia* 50 (4): 435–446. doi: 10.1016/j.neuropsychologia.2011.07.013.
A tutorial review that this author found to be very helpful for understanding the principles of multivariate encoding models.
- Squire, Larry R. 2006. "Lost Forever or Temporarily Misplaced? The Long Debate about the Nature of Memory Impairment." *Learning and Memory* 13 (5): 522–529. doi: 10.1101/lm.310306.
A detailed, authoritative review of consolidation block vs. retrieval-based accounts of anterograde amnesia.
- Yonelinas, Andrew P. 2001. "Components of Episodic Memory: The Contribution of Recollection and Familiarity." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 356 (1413): 1363–1374. doi: 10.1098/rstb.2001.0939.
Detailed theoretical and quantitative background for the ROC-based dual-process model of retrieval.

CHAPTER 12

SEMANTIC LONG-TERM MEMORY

KEY THEMES

- An intact MTL is not necessary for the preservation and retrieval of semantic memory.
- The study of patients showing (doubly dissociable) selective impairments at naming or identifying living things vs. nonliving things has given rise to three classes of theory: (1) knowledge is organized according to taxonomic categories (living vs. nonliving being the most salient); (2) knowledge is organized according to the functional modalities (e.g., visual perception of color and shape, motor sequences for manipulating tools) from which meaning derives; (3) lexical representation and access (i.e., words per se, not necessarily knowledge) is organized according to taxonomic categories.
- Important, and difficult, questions have been raised about the logical limits of what can be inferred from the results from single neuropsychological cases, as well as from demonstrations of double dissociations of function.
- The dependence on language of many experimental procedures used for testing semantic memory can complicate the interpretation of whether a particular set of neuroimaging findings reflects the processing of knowledge, per se, or reflects the language functions (often, lexical retrieval) engaged by the behavioral task.
- Attributing neuroimaging signal to knowledge vs. to visual perception and/or mental imagery can also be tricky.
- fMRI adaptation, which takes advantage of the phenomenon of repetition suppression, can be an effective technique for deducing key functional properties of a targeted region of cortex.
- Research on patients with different variants of frontotemporal dementia has been characterized by debates on whether the core deficit underlying systematic patterns of cognitive decline is semantic or language-based.
- Neuroimaging (both VBM and activation studies) and computational modeling indicate an important role for the anterior temporal lobe in the control, if not the storage itself, of semantic memory.

CONTENTS

KEY THEMES

KNOWLEDGE IN THE BRAIN

DEFINITIONS AND BASIC FACTS

CATEGORY-SPECIFIC DEFICITS

FOLLOWING BRAIN DAMAGE

Animacy, or function?

A PDP model of modality-specificity

The domain-specific knowledge hypothesis

THE NEUROIMAGING OF KNOWLEDGE

The meaning, and processing, of words

An aside about the role of language in semantics and the study of semantics

PET scanning of object knowledge

Knowledge retrieval or lexical access?

Knowledge retrieval or mental imagery?

THE PROGRESSIVE LOSS OF KNOWLEDGE

Primary Progressive Aphasia or Semantic Dementia, what's in a name?

Nonverbal deficits in fluent primary progressive aphasia?

The locus of damage in fluent primary progressive aphasia?

Distal effects of neurodegeneration

Entente cordiale

NUANCE AND CHALLENGES

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

KNOWLEDGE IN THE BRAIN

Questions about how we represent knowledge have been at the core of formal inquiry into human cognition, and the human condition, for as far back in history as we have written records. From the perspective of cognitive neuroscience, it's one of the domains in which there is considerable contact with scholarship in the humanities and the social sciences, including philosophy, linguistics, and cognitive development. It is also noteworthy for its relative lack of overlap with other branches of neuroscience, both because it is so difficult to study in nonhuman species, and because it deals with mental faculties that other animals simply may not possess.

As we have done in other chapters, we'll outline some of the key questions by considering findings from neuropsychology. This time, however, instead of starting with an exotic "new" neurological syndrome, we'll begin with our old friend H. M.

DEFINITIONS AND BASIC FACTS

Semantic memory is defined as memories for facts whose retrieval is not accompanied by information pertaining to how, when, or where that information was acquired – its retrieval is noetic. One of the striking findings with H. M. was that his memory for facts that he had learned prior to his surgery remained intact. So, too, did his vocabulary. Indeed, our knowledge of the meanings of words, often referred to as the **mental lexicon**, is an important subset of our semantic memory. Although it was initially believed that H. M.'s episodic memory for the distant past was intact, more recent interpretations, many of them prompted by the multiple trace theory, argue that even these were impoverished, "semanticized," and lacking the feeling of "reexperien[ing] ... something that has happened before." There is, however, no comparable debate about his semantic memory. For example, in 1953, H. M. was administered the Wechsler Adult Intelligence Scale (WAIS, i.e., an IQ test) as a part of preoperative cognitive testing in the run-up to his surgery. His performance on the subtests that measure lexical memory was within the normal range on that occasion, and it remained unchanged across

19 different testing sessions that were performed during the ensuing decades, through 2000. Thus, an intact MTL is not necessary for the preservation and retrieval of semantic memory.

How are semantic memories acquired? One formal account that we reviewed in *Chapter 10* is the slow cortical component of the complementary learning systems model of McClelland, McNaughton, and O'Reilly (1995). It is generally accepted that all memories begin as episodic, and gradually transition to becoming semantic. (This view, however, is not universally held, as we shall see in *Methodology Box 12.1*; but don't go there just yet.) A concrete example familiar to many readers might be one's knowledge of Barack Obama. The first time that a typical American (okay, let's pick on me again) may have heard of this individual may have been while Obama was campaigning to be President of the United States. Let's say that I first encountered the name while reading a newspaper article about the Democratic Primary race during the early months of 2008. (He was elected President in November 2008.) The next time I thought about Barack Obama may have been when a friend asked, "Which Democrat do you think will win the party's nomination?" To answer this question, I would have retrieved a memory of (the episode of) having read that newspaper article. Let's say that the third time I heard the name was while watching a televised debate between the candidates. At that point, would I think back to the newspaper article, or the discussion with the friend, each time the camera turned to Obama? Or would the knowledge *Black man, junior Senator from Illinois, Democratic candidate for President* just pop into my head noetically? If not by this third episode, shortly thereafter repeated iterations of this process would have created a representation that can be retrieved independently of any reference to any one of the episodes in which this information was encountered. In this way, the knowledge of "Barack Obama" becomes a *semantic* memory. Should I sustain bilateral damage to my MTL in, say, 2020 (pause here for a moment while author knocks on wood to negate this premonition), I would still be able to retrieve this knowledge after that time.

Another core fact about semantic memory is conveyed by its name: a semantic memory represents meaning. Further, we know that it is organized in

relation to its meaning. In the case of Barack Obama, for example, numerous experiments have demonstrated that if one is thinking about him, one can more quickly access information about John F. Kennedy (a US President from the 1960s) than if one was thinking of, say, Pelé, or John Lennon. Precisely how knowledge is organized, however, remains uncertain, and will be the question underlying much of the research that we'll review in this chapter.

And *now* we can turn to some exotic cases from cognitive neuropsychology.

CATEGORY-SPECIFIC DEFICITS FOLLOWING BRAIN DAMAGE

Animacy, or function?

Two seminal publications that have greatly influenced the contemporary study of the neural bases of semantics came from British neuropsychologist Elizabeth Warrington and her colleagues in the mid-1980s. First, Warrington and McCarthy (1983) described a patient (V. E. R.) whose large left hemisphere stroke produced a massive impairment in the ability to speak or to understand spoken language, including the following of simple instructions. Nevertheless, they were able to demonstrate some spared comprehension, in that, in response to a spoken word, she could point to a picture corresponding to the word. Her performance showed a noteworthy pattern, however, in that she produced many more correct responses when the words named flowers, animals, or foods than when they named manmade objects. Thus, she seemed to

have a relative impairment of knowledge for nonliving things. The next year, Warrington and Shallice (1984) reported four patients recovering from herpes encephalitis who showed the opposite pattern: all four demonstrated worse performance with living than with nonliving words. Further, two of these patients had sufficiently spared language abilities that they could be tested on picture naming, and here the pattern was the same. One of them (J. B. R.) named 90% of the nonliving objects he was shown, but only 6% of the living objects. The second (S. B. Y.) named 75% vs. 0%. Examples of S. B. Y.'s definitions given to spoken words, in *Table 12.1*, are particularly compelling, in that they illustrate that the deficit is not just a deficit of speech production, or general verbal fluency.

Together, these two studies produced a double dissociation, thereby seeming to rule out any trivial reasons that one category might be inherently more difficult than the other, and therefore more vulnerable to brain trauma of any kind. What remained, of course, was to explain this intriguing pattern of findings. One interpretation is that they provide evidence that knowledge is represented according to taxonomic categories in the brain. This view has been championed perhaps most vigorously by Italian-born, US- and Italy-based cognitive neuropsychologist Alfonso Caramazza. In one formulation of a “domain-specific knowledge hypothesis,” for example, he and Shelton (1998) propose that “evolutionary pressures have resulted in specialized mechanisms for perceptually *and* conceptually distinguishing animate and inanimate kinds ... leading to a *categorical* organization of this knowledge in the brain.” We will return to this hypothesis in a bit, but first let's consider an alternative,

TABLE 12.1 Example definitions to spoken words by patient S. B. Y.

<i>Duck</i> : an animal	<i>Wheelbarrow</i> : object used by people to take material about
<i>Wasp</i> : bird that flies	<i>Towel</i> : material used to dry people
<i>Crocus</i> : rubbish material	<i>Pram</i> : used to carry people, with wheels and a thing to sit on
<i>Spider</i> : a person looking for things, he was a spider for his nation or country	<i>Submarine</i> : ship that goes underneath the sea

Source: Warrington and Shallice, 1984. Reproduced with permission of Oxford University Press.

Tip 12.1

Remarkably, Warrington and McCarthy (1987) later reported a patient with a mirror-opposite profile: general impaired performance with nonliving things but spared performance with metals, types of cloth, musical instruments, and precious stones!

because one of the factors that makes the domain-specific knowledge hypothesis provocative is that many investigators, including Warrington herself, have *not* interpreted these patterns of category-specific impairment as evidence for categorical organization in the brain.

Warrington and Shallice (1984) noted that patient J. B. R., for example, also showed impaired performance for several categories of nonliving objects: metals, types of cloth, musical instruments, and precious stones (*Tip 12.1*). Thus, perhaps the (relatively) selective impairments of knowledge for living vs. nonliving things in these patients were byproducts of two different underlying factors with which the categories of living and nonliving are highly correlated. Warrington and Shallice proposed that these factors might be two cardinal functions of the brain, perception and action. We primarily experience living things, they reasoned, by vision, whereas we experience objects by how we use them. Thus, patterns of relative impairment or sparing of knowledge about living or nonliving things may result from the damage to brain areas primarily involved in visual perception or in action planning. This alternative can be called a modality-specific scheme, because it posits that the organizing factor for semantics is the functional modality (e.g., vision, motor control) that a given piece of knowledge relies on.

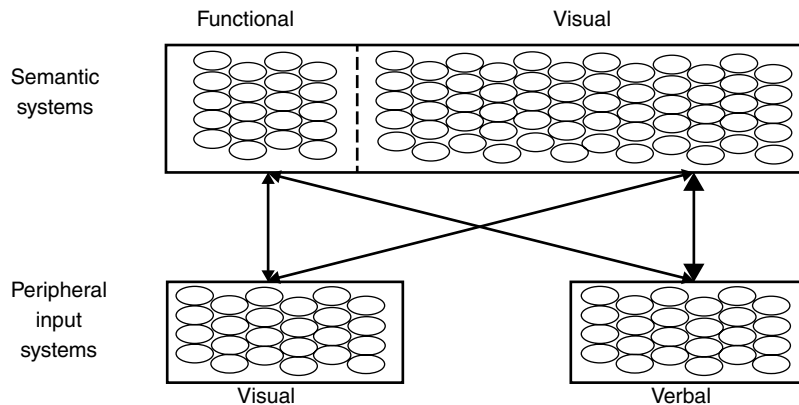
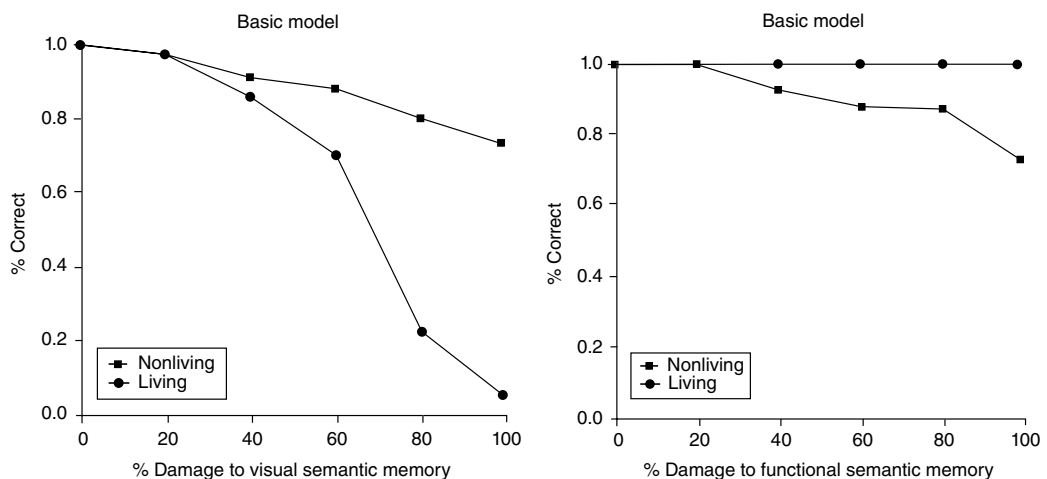
A PDP model of modality-specificity

Martha Farah (who we met in *Chapter 9*) and James McClelland (*Chapter 10*) sought to capture this intuition with the PDP model that is illustrated in *Figure 12.1.A*. First, to quantify the idea that living things are more strongly associated with visual representations and nonliving things with their function, they obtained dictionary definitions for each of the living and

nonliving words used by Warrington and Shallice (1984), asked 21 subjects to go through each definition and to underline “all occurrences of words describing any aspect of the visual appearance of an item,” and asked 21 others to underline “all occurrences of words describing what the item does or what it is for” (1991, p. 342). The results identified an average of 2.68 visual descriptors for living things and 1.57 for nonliving things, and an average of 0.35 functional descriptors for living things and 1.11 for nonliving things. Thus, a visual:functional ratio of 7.7:1 for living things and 1.4:1 for nonliving things. (Note that the ratio for nonliving things was not <1 , as Warrington and Shallice’s explanation assumed.) These ratios were then used to determine the proportion of units in the semantic layer of the PDP model that were labeled “visual” vs. “functional,” as well as the number of units of each type used to represent each item.

The model represented 10 living and 10 nonliving things, each as a distinct, randomly determined pattern of +1 and −1, with each item represented by the full set of 24 visual input units and 24 verbal input units, but each living thing represented by an average of only 16.1 visual and 2.1 functional units in the semantic layer, and each nonliving thing by an average of only 9.4 visual and 6.7 functional units in the semantic layer. This captured the ratios derived from the dictionary definitions. Note that nowhere in the model was “livingness” or “nonlivingness” explicitly represented.

Next, the model was trained, by iteratively adjusting weights in the manner described in *Research Spotlight 9.1*, until, upon being given each visual item, it produced the correct pattern in the semantic and verbal layers, and upon being given each verbal item, it produced the correct pattern in the semantic and visual layers. Finally, it was time to simulate brain damage. As illustrated in *Figure 12.1.B*, this was done by progressively turning off 20%, 40%, 60%, 80%, or 100% of the units in the visual partition of the semantic layer (left plot) or in the functional partition of the semantic layer. The results produced the double dissociation that has been observed across cases, in that progressively greater damage to visual semantic memory preferentially impaired performance with living things, and progressively greater damage to functional semantic memory preferentially impaired performance with nonliving things. Like any other

FIGURE 12.1 Modeling category specificity as emergent from a modality-specific computational architecture.**FIGURE 12.1.A** Schematic diagram of the model of Farah and McClelland (1991), as rendered by Rogers and Plaut (2002). Source: Rogers and Plaut, 2002. Reproduced with permission of Psychology Press.**FIGURE 12.1.B** Performance of the model as a function of the percentage of damage to visual semantic memory (top plot) or to functional semantic memory (bottom plot). Source: Farah and McClelland, 1991. Reproduced with permission of APA.

computational simulation, these results of the modeling of Farah and McClelland (1991) can't prove that knowledge is organized in a modality-specific rather than a category-specific manner. It does, however, provide a quantitatively explicit proof of concept for an alternative that many would argue is more neurally plausible. For example, in a subsequent paper, Farah, Meyer, and McMullen (1996) opined that a drawback of category-related models is that they "represent a significant departure from everything else we know

about brain organization, according to which subsystems are delineated by function or modality but not by semantic content" (p. 143).

The domain-specific knowledge hypothesis

To defend the category-based interpretation of category-specific deficits, Caramazza and colleagues rely heavily on logical argumentation, and on additional case studies. From the latter, for example, Caramazza and Shelton (1998) observe that the categories of

animals, fruits and vegetables, and artifacts can be impaired in isolation across patients. From this observation they have proposed that “these, and only these, three categories form the basis for the organization of conceptual knowledge” (p. 19). The basis for these categories would be “evolutionary pressures [that] led to specific adaptations for recognizing and responding to animal and plant life” (p. 20). The argument is plausible. One can certainly see how the ability to quickly recognize and categorize an animal as potential predator or prey, or a plant as potential food or medicine or poison, might raise the odds of survival and, therefore, to evolutionary selection of those traits. However, although such ideas about evolutionary factors can be used to assemble a plausible, if not intriguing, “back story” for a model, such ideas are impossible to test, and so offer limited explanatory power. Further, because they are necessarily post hoc, one quickly finds that analogous, and equally plausible, evolutionary accounts can be marshaled for alternative models. Thus, for example, an equally plausible story can be told about the evolution of sensory and motor systems, and how increased computing power of the evolving primate brain led to the abstraction and elaboration of portions of these two core systems into the semantic systems that humans enjoy today. As there is no direct way to test which of these evolutionary stories may be a closer approximation of the truth, neither can contribute directly to a mechanistic understanding of how this aspect of cognition is carried out by the brain.

The ideas considered in the previous paragraph illustrate one complication of working with neuropsychological data, particularly with single case studies. Without question, neuropsychology is an invaluable tool for understanding human behavior, which is complex, variable, and noisy. Systematic deviations from typical patterns of behavior, particularly when they can be associated with specific patterns of brain damage, can provide important insight into the organization and principles of a particular aspect of cognition. To draw on just one familiar example, the case of H. M. provided evidence for a fundamental distinction between processes associated with LTM encoding vs. with LTM storage and retrieval. Importantly, however, it was the subsequent replication of these findings with other patients that provided confirmation of the generalizability of what

was found in one patient to principles about how all human minds and brains are organized. This sets up a tension – a conceptual tension as well as passionate debates among scientists – about how to weight the single case study. This, and other complications associated with the interpretation of neuropsychological data, are considered in *Methodology Box 12.1*.

In view of these complications associated with cognitive neuropsychology, it is not surprising that semantic memory researchers have turned to functional neuroimaging as an important source of data. So too, now, will we.

THE NEUROIMAGING OF KNOWLEDGE

The meaning, and processing, of words

Many pioneering neuroimaging studies addressing the questions that we have reviewed up to this point have come from Alex Martin and his collaborators at the National Institutes of Health (NIH). Martin trained as a neuropsychologist, and spent the early part of his career studying the breakdown of language and memory functions in patients with Alzheimer’s disease and HIV-AIDS. In 1989 he moved to the National Institute of Mental Health (NIMH), where he quickly fell in with Haxby, Ungerleider, and colleagues and soon began adapting their neuroimaging methods for the study of semantic memory. As a starting point, Martin considered the PET study of Petersen, Fox, Posner, Mintun, and Raichle (1988; *Figure 6.5.A* in this book), specifically the contrast that isolated verb generation and is entitled “Thinking about words.” The fact that this contrast revealed greater activity in left inferior PFC (centered in BA 47) for verb generation than for either reading visually presented words or repeating auditorily presented words suggested a role for left PFC in semantic processing, independent of input modality. That is, this finding made left inferior PFC a candidate locus for an amodal semantic store. However, Martin doubted that semantics would be localizable to this particular region of the brain, for a number of reasons. First, theories from cognitive psychology and from linguistics posited that “the meaning

METHODOLOGY BOX

12.1 Problems in neuropsychology: How definitive is a single case study? A double dissociation?

We have seen that many “firsts” were discovered with the patient H. M. In retrospect, however, the enduring contributions from this research are those facts that have been replicated in other patients, confirmed via complementary studies with other species, refined via computational and neuroimaging research, etc. (Indeed, in many cases the initial “facts” deduced from patterns in H. M.’s behavior (such as the idea of completely spared episodic memory for events experienced prior to the penumbra of retrograde amnesia) have been revised in light of subsequent findings and theoretical developments (for example, multiple trace theory and associated research).) How, then, are we to think of other single cases as they are discovered? Which may be “the next H. M.,” in the sense that they presage a series of discoveries and replications that will generalize to human cognition? (After all, “someone has to be first.”) Alternatively, which individual cases will turn out to be single, idiosyncratic “outliers” that present with peculiar patterns of results that, in retrospect, turn out to never be replicated or extended? One example of the latter might be the patient K. C., who has been described in several studies as having a selective deficit in the encoding of episodic, but not semantic, LTM. A face-value acceptance of the findings from K. C. cannot be reconciled with the “generally accepted” view that “all memories begin as episodic, and gradually transition to becoming semantic.” Although Tulving – who unarguably occupies the highest echelon of respected and influential theoreticians of memory – has made this argument very cogently (for review, see Rosenbaum et al., 2005), the absence, in the years since the initial descriptions of K. C. in the 1980s, of any additional patients presenting a similar behavioral profile, may explain the fact that this case has had negligible influence on most current views of LTM encoding. Whether or not the same is (or, will be) true of single cases in the literature on the organization of semantic memory remains a matter of vigorous debate (with, e.g., Mahon and Carramazza (2009) arguing for the importance of single cases, and Rogers et al. (2006) calling their importance into question).

Often tied up in these debates is the more technical question of *How important is the double dissociation formed by comparing case A with case B?* This question has arisen from neural network simulations demonstrating that double dissociations can be procured from unitary computational architectures. Although the details of these arguments are beyond the scope of what we can cover in this textbook, a feel for the essence of some of these arguments is captured in a chapter from Rogers and Plaut (2002):

apparent double dissociations even in ... homogeneous network[s] with no assumed neuroanatomic specialisation ... [and] double-dissociations ... aris[ing] from damage to anatomically distinct areas that are in no way specialised to subserve the cognitive functions dissociated ... call into question the conclusion ... that the dissociated functions must be subserved by modules that may be damaged independently. (p. 252)

Counterarguments have pointed to simulations such as Farah and McClelland (1991), in which meaningful structure can exist, and can therefore be selectively impaired, within “single” computational architectures.

of a concrete noun is not unitary, but rather is composed of parts – specifically, knowledge about the physical and functional properties of the object.” One would therefore predict word meanings to be anatomically distributed. Complementary to this perspective were the computational and neural canons that we have established in the previous two chapters, that long-term memories are comprised of the association of the neural representations of many sensations, actions, emotions, and so on. Additionally, he reasoned that “even if semantic networks were confined to a single region, the neuropsychological evidence suggested that the critical area would be the left temporal, not the left frontal, lobe” (Martin, Ungerleider, and Haxby, 2000, pp. 1023–1024).

An aside about the role of language in semantics and the study of semantics

Note that, for each type of study that we’ve considered thus far in this chapter, language has been central to the theoretical motivation and/or the experimental procedures employed. Warrington and McCarthy’s 1983 paper is entitled “Category Specific Access Dysphasia,” **dysphasia** being an impairment in the ability to communicate following brain damage. Thus, their title appeals to the directly observable clinical presentation of the patient, rather than to the psychological construct of semantic memory. Indeed, experiment 1 from this paper reported the phenomenon that they presumably set out to study in this patient, a dissociation between her relatively spared ability to match pictures of common objects to exemplars of the objects themselves vs. a relative impairment in matching spoken words to the same objects. The subsequent several experiments then examined such factors as how the rate of presentation affected her performance, the effect of using spoken vs. written cues, and so on. It was only at experiment 10 (sic!) that the issue that this paper is famous for arose: “An incidental observation in a pilot study ... suggested that V. E. R.’s comprehension deficits might to some extent be more severe for some taxonomic categories than others” (p. 870.). The Farah and McClelland (1991) model, intended to emulate the Warrington studies, was trained, and tested, by mapping verbal labels to pictures, and vice versa. Peterson et al. (1988) studied the generation of a verb from a visually or

auditorily presented noun, and many, many subsequent neuroimaging studies have also used linguistic material to study semantics. Perhaps this is a necessary association? Whether language is fundamental to the mental representation of knowledge, or at least to its acquisition, is a question that has occupied philosophers for centuries, and is certainly not one that we can hope to resolve in this textbook. What is important for our purposes is to be aware of this association while evaluating research that we’ll review here, and of the competing claims about the centrality of language to the impairments seen in various neurological syndromes, and to patterns of activity observed in brain imaging studies.

PET scanning of object knowledge

Alex Martin’s analysis of the neuropsychological literature drew heavily on the visual agnosias, which could produce selective impairment of particular stimulus attributes, such as cortical color blindness (achromatopsia) and motion blindness (akinetopsia). Thus, his first PET study, published in 1995, was designed to isolate the contributions to object knowledge of the cortical representation of the features *color* and *action*. It presented subjects with black-and-white line drawings of common objects, and in one scan asked subjects to name them, in a second, to name a color associated with that object, and in a third, to name an action that could be performed with that object. With a procedure comparable to that used by Petersen et al. (1988), the data were analyzed by subtracting the object-naming images from both the color- and action-naming images, thereby emphasizing the processing of those two object attributes. The results, presented in *Figure 12.2*, indicated that color naming was selectively associated with activity in a region of ventral temporal cortex, just anterior to a region of ventral occipital cortex associated with color perception, and that action naming preferentially activated a broad expanse of left lateral cortex encompassing posterior middle temporal gyrus (MTG), superior temporal gyrus (STG), and posterior inferior parietal lobule (IPL). In addition to some overlap in these posterior regions, both tasks also preferentially activated a large region of left PFC and premotor cortex (PMC), with the action-generation task

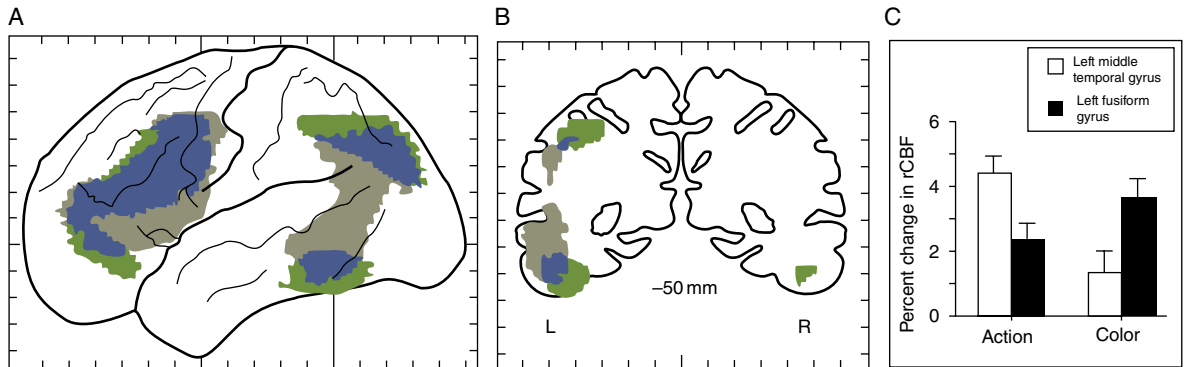


FIGURE 12.2 Superimposed subtraction images of object attribute naming from black-and-white line drawings. Green = object color generation (minus object naming); gray = object action generation (minus object naming); and blue indicates the overlap. **B.** Coronal slice is 50 mm posterior to the “0 mm” location, along the rostrocaudal axis of the anterior commissure. **C.** Peak magnitude of these effects in two temporal lobe regions. Source: Martin, Haxby, Lalonde, Wiggs, and Ungerleider, 1995. Reproduced with permission of AAAS.

preferentially engaging posterior inferior frontal gyrus, including Broca’s area. In a second experiment, subjects performed the same tasks, but in response to written words instead of line drawings, and the results were, by and large, the same.

Martin et al. (1995) interpreted their results as evidence that “knowledge of colors and of actions is represented in discrete cortical areas ... [that are] proximal to color and motion perceptual-processing areas” (p. 104). They further argued for a framework in which the representation of knowledge about objects is distributed among the representation of the features that make up an object, and tied to the modality that characterizes these features (e.g., the perception of chrominance and the perception of motion). Note the similarities between this perspective and the modality-specific models of Warrington and colleagues and of Farah and McClelland (1991). Indeed, consistent with the modality-specific view were the results of a second study that contrasted the silent naming of animals with the silent naming of tools (Martin, Wiggs, Ungerleider, and Haxby, 1996), effectively a PET version of the Warrington studies. The findings from this study were also consistent with a modality-specific model: naming animals was associated with greater activity in visual cortex, and naming tools with greater activity in regions that largely overlapped those identified with action generation in the 1995 study – posterior middle temporal gyrus and premotor cortex (Figure 12.3). In

interpreting these results, and in several subsequent studies, Martin and colleagues have argued that the tool-related activity in the posterior middle temporal gyrus relates to the processing of artifact motion, and that the tool-related activity in premotor cortex corresponds to movement plans associated with object use. The involvement of premotor cortex, in particular, has prompted the suggestion of a role for mirror neurons in the representation of tool semantics. There are also clear points of contact with the rapidly growing “embodied cognition” perspective in psychology.

To just review the results from these two PET studies as we have done up to this point, and to then move on to the next topic, would be to misrepresent this field of study. There are many conceptual nuances, and pragmatic challenges, with which one must grapple when studying a cognitive faculty as abstract as *knowledge* with a method as concrete as functional neuroimaging. Publishing in this arena often sparks vociferous debate. To get a feel for some of this nuance and challenge, let’s follow the trajectory, as it were, of just one claim from the Martin et al. (1995) study, considering some of the scrutiny it has received and how it has, in turn, been critiqued and defended. The claim that we’ll follow is that activity in left fusiform gyrus is associated with the storage of color-related knowledge about objects (Figure 12.2). We’ll begin by returning to the Martin et al. (1995) paper itself, in whose pages we can already find evidence for objections to this claim.

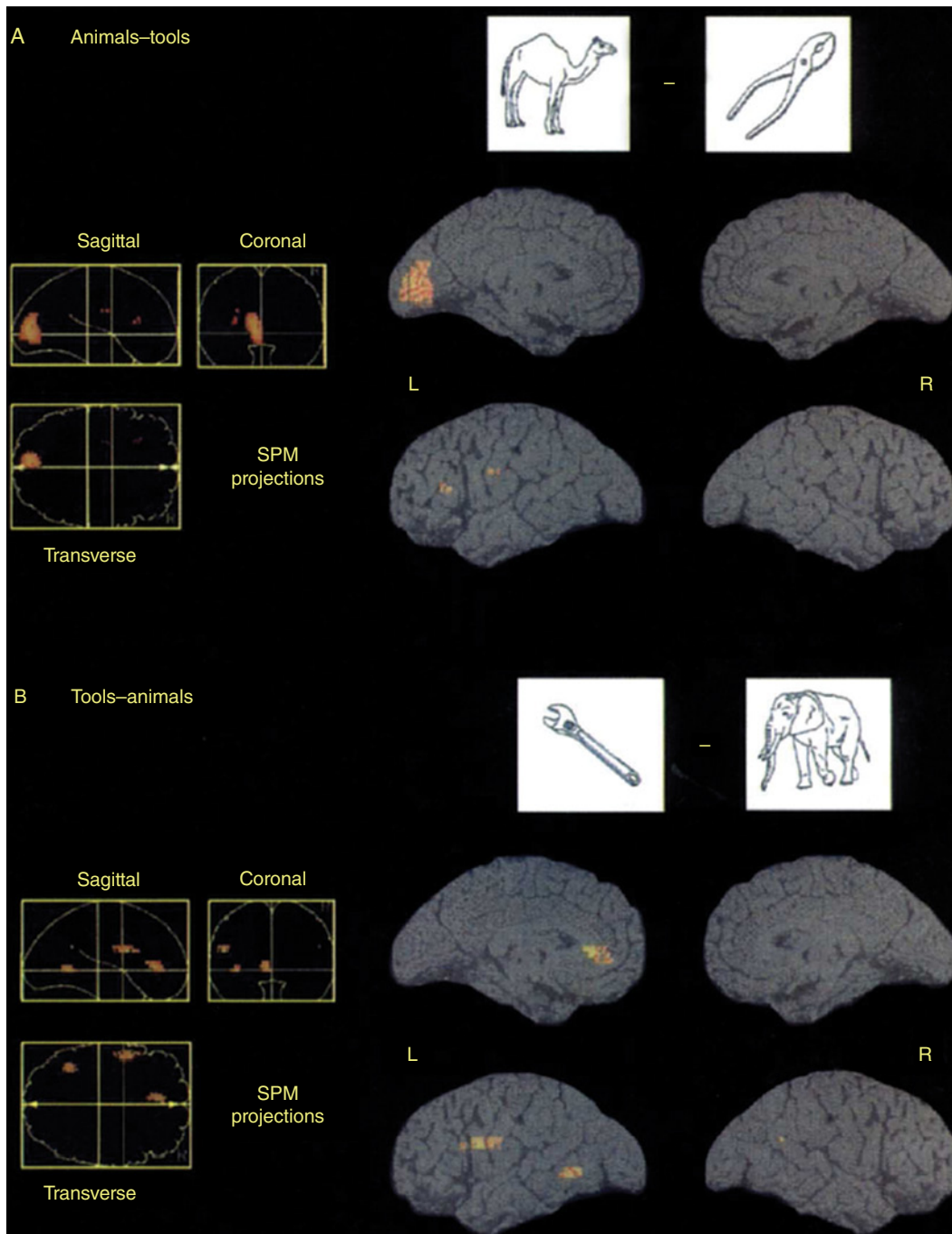


FIGURE 12.3 PET activity is greater for silently naming animals than silently naming tools (**A**) prominently in left V1 and adjacent extrastriate cortex, and the converse (**B**) in left posterior middle temporal gyrus, left anterior cingulate cortex, and left premotor cortex. Source: Martin, Haxby, Lalonde, Wiggs, and Ungerleider, 1995. Reproduced with permission of Nature Publishing Group.

Knowledge retrieval or lexical access?

In the discussion of their findings, Martin and colleagues (1995) include a footnote to reinforce the interpretation of their findings in terms of “knowledge,” *per se*, rather than of activation of the mental lexicon:

Findings from studies of patients with focal brain lesions argue against alternative interpretations of these activations as the sites for storage of the color and action words themselves, or of word-specific retrieval mechanisms located at sites distant from where information is stored. A patient with a color anomia can produce color words. What the patient cannot do is correctly name the colors seen or correctly answer questions about object-associated colors (the patient will respond with an inappropriate color word). To show areas of activity by functional brain imaging technologies, it is necessary for subjects to be engaged in performing a task. Thus, it might be formally more accurate to state that the activations in the temporal lobes show the sites of storage during the act of retrieval. In this sense, however, storage and retrieval cannot be distinguished. (p. 105)

Although one can't know for sure (and I am deliberately choosing not to contact Alex Martin to confirm or deny my speculation), it is not infrequently the case that a footnote such as this gets added to a manuscript after it has gone through peer review. For example, if there's a particular point about which the author and a reviewer disagree, sometimes an editor will say something like this to the authors: “Although you disagree with this argument raised by the reviewer, you need to explicitly raise it in your manuscript. In addition, if you like, you can also give your rationale for disagreeing with it.” Whether or not a similar scenario played out in this instance, what is clear is that this footnote is intended to refute an alternative interpretation with which the authors strongly disagree. There are several instructive aspects to this footnote. First, it highlights the fact that, particularly in this research area, the design and interpretation of neuroimaging studies is heavily influenced by patient studies, and vice versa. Second, it highlights one of the pragmatic challenges raised by the tight coupling of language and semantics. Finally, it highlights a second alternative perspective (in addition to the domain-specificity model) to the modality-specific model: the true driver of patterns of neural

activity in the left temporal lobe, such as those described here, and, by extension, of category-specific deficits in patients with lesions to these areas, may be lexical retrieval.

The lexical-retrieval view has been advocated by the (Portuguese-born, US-based) husband-and-wife team of neurologist Hanna Damasio and neurologist Antonio Damasio. They have argued, from patient and neuroimaging data presented in a 1996 paper, that lateral temporal-lobe regions, although not classical language areas, make important contributions to the retrieval of words, and do so in a domain-segregated way, with lexical access to the names of people mediated by the temporal pole, to the names of animals by mid-inferior temporal gyrus wrapping onto the ventral surface of the temporal lobe, and to the names of tools by posterior middle temporal gyrus. They have interpreted the core deficit of patients with damage to each of these regions to be one of lexical retrieval (i.e., word-finding), because the naming deficits of their patients were demonstrated after it had been determined that they understood the concept of the item in the picture. (This specific claim, incidentally, is one with which Caramazza and Shelton (1998) take issue.) Thus, Damasio, Grabowski, Tranel, Hichwa, and Damasio (1996) have argued for a different functional interpretation of the ventral temporal-lobe activity that Martin et al. (1995) attribute to color knowledge (as well as for a different functional interpretation of the area that Martin et al. (1995) attribute to action generation).

Recall that we previously noted that, in composing their title, Warrington and McCarthy (1983) emphasized the directly observable clinical presentation of their patient (“dysphasia”) rather than that psychological construct that the patient's clinical presentation implicated. It may be that the emphasis that the Damasios have placed on a lexical explanation for their data also derives, at least in part, from their experience in the clinic. Both are behavioral neurologists whose contributions to cognitive neuroscience have been strongly anchored by their experience and expertise in working with patients. Further along in this chapter we'll also see that a naming interpretation has been favored by many neurologists in relation to a neurodegenerative disease that they have referred to as “primary progressive aphasia” ([aphasia](#) being an ina-

bility to use language), whereas many approaching the same disease from a psychological perspective call it “semantic dementia.”

Up to this point, it has been patients from the neurology clinic (victims of stroke or of herpes encephalitis) who have featured most prominently in the literature

that we have considered. *Real-World Applications 12.1* introduces a quite different group of individuals – synesthetes – and summarizes findings from a PET imaging study of synesthesia on which Martin has drawn to rebut lexical-access interpretations of the color-naming data.

REAL-WORLD APPLICATIONS

12.1 Synesthesia

Synesthesia (constructed from the Greek *syn* (together) *esthesis* (sensation)) is a condition in which the sensation (and/or mental imagery) of certain stimuli leads to the conjoint perceptual experience of the sensed stimulus plus an additional component from a different sensory modality or domain of information, despite the additional component not being physically present. Examples of cross-modality synesthesia include certain sounds evoking the perception of certain colors, or certain tastes evoking the perception of certain shapes. An example of cross-domain synesthesia is grapheme–color synesthesia, in which certain letters and/or digits are always perceived as appearing in a particular color, regardless of the true chrominance of the medium from which they are formed. (In the latter case, the individual can typically discern both the true color and the synesthetic color, as though the latter is a translucent “cloud” floating above it.) The causes of synesthesia aren’t fully understood, although it has a heritable component. This has given rise to the idea that synesthesia arises from a mutation that produces a deviation from normal patterns of pruning excess neural connections during development, such that connections remain among otherwise unassociated sensory regions. (One intriguing idea is that all infants are born synesthetes, but most lose it as perception is refined by experience.)

This feature box could have been a *Research Spotlight* or a *Thought Box*, but I opted for *Real-World Applications* because, although this sounds like an extraordinary condition, its prevalence is estimated to be as high as 4.4% of the population (*Tip 12.2*). Clinically, synesthesia is not known to be comorbid with any other neurological or psychiatric condition, and so synesthetes are otherwise cognitively typical. From the standpoint of a neuropsychologist, this means that the investigator can draw on the patient’s introspection for insight into the condition. (The same is certainly not true of patients whose language abilities have been compromised by a stroke, or who are suffering from a dementing neurodegenerative disease.)

The relevance of synesthesia for this chapter is that a PET scan of synesthetes has provided independent corroboration of the assertion by Martin et al. (1995) that ventral temporal lobe activity in response to color naming corresponds to the perception of color, to the meaning of a color, rather than to the activation of color words. The study was conducted by Italian neurologist Eraldo Paulesu, at the time a research fellow at the MRC Cyclotron (i.e., “PET scanning”) Unit in London, and his colleagues. They scanned six word–color synesthetes listening to spoken words known to be synesthetic for each of them, and listening to tones. Words, but not tones, triggered the perception of color in the synesthetes. In synesthetes and control subjects, word stimulation produced greater activity than tone stimulation in classical language areas of left perisylvian cortex (*Chapter 18*). In synesthetes, but not controls, words also produced greater activity in several posterior regions, including the ventral temporal region selectively activated by Martin et al. (1995) in the color-naming condition. The relevance for the Martin et al. (1995) study is that, for these individuals, we know that they experience colors when they hear these words, not “just” the lexical representations of these colors.

Tip 12.2

Many synesthetes seem to go through life without realizing that this aspect of their conscious experience is not “normal.” Anecdotally, I have heard more than one colleague recount how a student has presented him- or herself to the professor after a lecture on synesthesia, the student not realizing until just then that his/her synesthesia was out of the ordinary.

Knowledge retrieval or mental imagery?

Yet another possible problem with Martin’s interpretation of his 1995 results was that the regions highlighted by his results are associated with mental imagery (e.g., imagining the color of a hammer or imagining someone swinging it). From this perspective, the temporal-lobe activity elicited in the Martin et al. (1995) study may have had little to do with conceptual processing, *per se*. Here’s how he summarized the argument in a later review paper: “... posterior temporal activity [may] simply reflect the explicit retrieval of visual object imagery that accompanies task performance ... [By this account,] stimuli denoting concrete objects, either pictures or words, trigger visual imagery, which then recruits posterior ventral temporal cortex indirectly; this area does not do any conceptual work.” (Martin, 2007, p. 24) An argument against this alternative account of the color-naming finding comes from studies leveraging the phenomenon of “fMRI adaptation” (a.k.a., “repetition effects”) which is described in *Methodology Box 12.2*.

The study by Simons et al. (2003) that is described in *Methodology Box 12.2* and illustrated in *Figure 12.4* is one of several senior-authored by Daniel Schacter, a cognitive psychologist who trained with Endel Tulving and, during the 1980s and 1990s, came to be known as arguably the world’s foremost authority on repetition priming. Initially in collaboration with Randy Buckner, Schacter and his mentees performed many of the first fMRI studies of this phenomenon. Among them, this study by then-visiting scholar Jon Simons and colleagues (2003) was one to which Martin could refer as evidence against the possibility that the left fusiform gyrus activity produced by his color-naming condition corresponded to “just” visual imagery that did not “do any conceptual work.”

Right. So that wraps things up for our “case study” of the color-naming claim. The realm of nuance and challenge to which we will devote the remaining portion of this chapter is that of a neurodegenerative disease whose cognitive impairments manifest prominently in language and semantic memory.

THE PROGRESSIVE LOSS OF KNOWLEDGE

Unlike the acute-onset neurological syndromes that we have considered up until now (stroke, viral encephalitis, and, indeed, surgical excision), neurodegenerative diseases are characterized by progressive degeneration of the integrity of the affected neural systems, and in the behavior(s) that depend(s) on those systems. *Dementia* refers to the progressive impairment of intellect and behavior that, over time, dramatically alters and restricts normal activities of daily living. Most dementias are of “insidious onset,” medicalese for “subtle and gradual.” The most common form of dementia is Alzheimer’s disease, which typically initially manifests as a progressively worsening anterograde amnesia, and gradually compromises other mnemonic and other cognitive functions. Atypical dementias are a class that first manifest in domains of cognition other than episodic memory. Of interest for our present purposes is a subtype of frontotemporal dementia that has variously been called *primary progressive aphasia* or *semantic dementia*.

Primary Progressive Aphasia or Semantic Dementia, what’s in a name?

As alluded to previously, there was a period during which one’s preferred label for this disease seemed to reflect, in part, the scholarly tradition from which one approached it. In the 1970s, neuropsychologists Elizabeth Warrington in the UK and Myrna Schwartz and her colleagues in Philadelphia both described cases of dementia with **anomia** (naming difficulty) that seemed to reveal a selective impairment of semantic memory. Subsequently, neurologists who saw similar patients in clinic labeled the condition semantic dementia. In parallel, in 1982, the behavioral neurologist M.-Marsel Mesulam wrote a highly

METHODOLOGY BOX

12.2 Repetition effects and fMRI adaptation

Recall, from *Chapter 10* and *Chapter 11*, that repetition priming is a type of nondeclarative memory in which previous exposure to a stimulus biases or facilitates the processing of that stimulus (or fragments thereof) when presented on subsequent occasions. At the neural level, a correlate of repetition priming is the repetition suppression effect, whereby neurons that responded to the initial presentation of the stimulus with an increase in firing rate respond with a smaller increase to the second and subsequent presentations. With neuroimaging, an analogous effect was shown by Buckner and his mentors at Washington University with PET studies of word-stem completion, and later, more definitively, by Buckner, Dale, and colleagues at the Massachusetts General Hospital-NMR Center with their rapid event-related fMRI technique.

A methodological innovation that takes advantage of this phenomenon, called **fMRI adaptation**, was developed by the Israeli team of Kalanit Grill-Spector and her mentor Rafael Malach. To deduce the functional properties of a region of visual cortex, one first presents a stimulus several times to produce a repetition-suppression effect. Then, one begins varying the stimulus properties along a single dimension. If the region of interest is invariant to changes along that dimension, its response will remain suppressed. If, however, it is sensitive to changes along this dimension, any change will result in a “release from adaptation” and a return to its initial level of response.

Illustrated in *Figure 12.4* is a study that employed the logic of fMRI adaptation. Subjects were presented with images of familiar objects (both living and nonliving) and asked to indicate for each, with a button press, whether or not it would fit into a shoe box. During the study phase, each of 32 items was presented on three occasions. During the test phase (the transition from study to test occurring without the subjects’ knowledge), each of 64 items was presented once. Of these, 32 were repeats of the 32 study items, 16 were different exemplars of the same kind of item as the 32 (“Different”), and 16 were items that had not appeared among the 32 studied items (“Novel”).

As illustrated in *Figure 12.4*, three of the regions of interest in this study showed an identical response profile: suppressed activity for a repeated item, but a release from adaptation for a same-type-but-different-token (i.e., Different) item. These regions, therefore, were assumed to encode the visual characteristics of stimuli, but not more abstract properties. (E.g., for a Study picture of an octopus, if a region’s response to a physically different picture of an octopus is not suppressed, it must be that this region doesn’t register that the two physically different images correspond to the same kind of thing. The interpretation is that this region doesn’t encode the meaning of “octopus-ness.”) The left fusiform gyrus ROI, in contrast, showed only a partial release from adaptation. (i.e., its response to a Different stimulus showed a level of adaptation intermediate between a repeated image and a Novel image.) Thus, this region’s activity registered the semantic similarity between the two images, even though their physical properties differed. (More detail at *Web Clip 12.1*.)

Although repetition effects have been fairly well characterized at the neural level, it is uncertain how fMRI repetition/adaptation effects are brought about. A paper from Grill-Spector, Henson, and Martin (2006) considers three possible mechanisms: fatigue, sharpening, and facilitation.

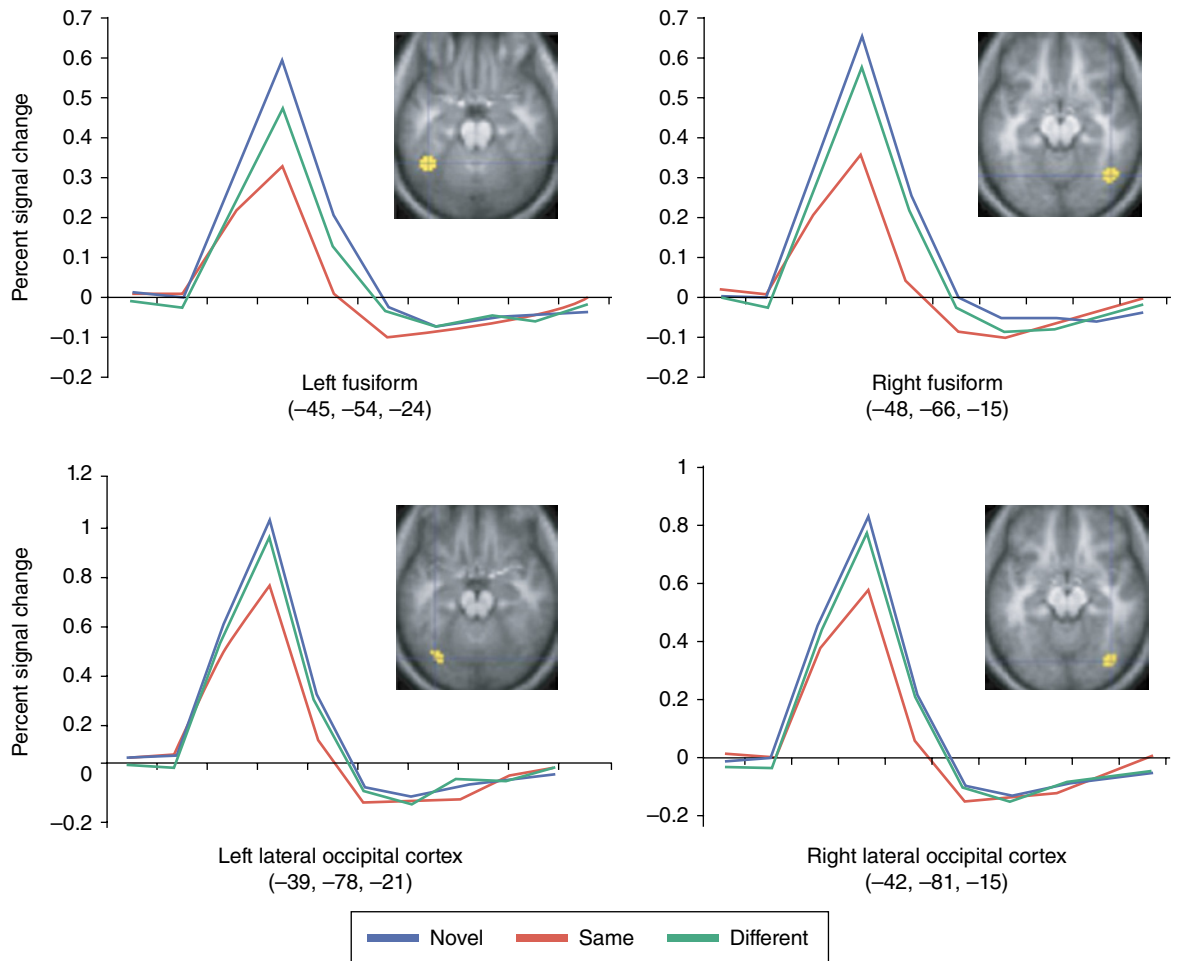


FIGURE 12.4 Repetition effects in four posterior regions. Each plot shows the fMRI stimulus-evoked response to the presentation of a picture that is either being seen for the first time (“Novel”), is a repeat presentation of a stimulus that had already been presented on three previous occasions during the same scan (“Same”), or is a different exemplar of one of the already-thrice-presented items (“Different”). Source: Adapted from Simons, Koutstaal, Prince, Wagner, and Schacter, 2003. Reproduced with permission of Elsevier.

influential paper describing patients with a condition he labeled “primary progressive aphasia,” and has subsequently characterized it as follows:

Patients with Alzheimer’s disease come to medical attention because of forgetfulness, usually accompanied by apathy. Misplacing personal objects, repeating questions, and forgetting recent events are among the presenting symptoms. Although the patient may forget people’s names, word-finding during conversation is usually not

a major problem. In contrast, patients with primary progressive aphasia come to medical attention because of the onset of word-finding difficulties, abnormal speech patterns, and prominent spelling errors. Primary progressive aphasia is diagnosed when other mental faculties, such as memory of daily events, visual and spatial skills (assessed by tests of drawing and face recognition), and behavior (assessed by a history obtained from a third party), remain relatively intact; when language is the only area of prominent dysfunction for at least the first

two years of the disease; and when structural brain imaging studies do not reveal a specific lesion, other than atrophy, that can account for the language deficit. (Mesulam, 2003, pp. 1535–1536)

In this and a contemporaneous review, Mesulam and a formidable line-up of colleagues asserted (via the title of their paper) that the language impairment is the core deficit in all forms of primary progressive aphasia, and poor word comprehension a peripheral “halo” in one subtype that they termed the fluent subtype of primary progressive aphasia. Alternatively, they argued that there is a second distinct neurodegenerative syndrome characterized by two independent factors: an aphasia due to pathology in left hemisphere perisylvian language cortex; and an associative agnosia due to pathology in the inferior temporal cortex.

In the remainder of this chapter we will consider a series of studies advocating an alternative view, whereby the core deficit of the putatively fluent variant of primary progressive aphasia is, in fact, semantic memory. The pedagogical strategy here is analogous to what we did with the color-naming finding from Martin et al. (1995) – by zooming in for a fine-grained consideration of this particular debate we will encounter experimental methods and lines of argumentation that illustrate broader themes about how scientists think about the neural bases of semantic memory. Much of this work that we will consider has been led by neurologist John Hodges and neuropsychologist Karalyn Patterson, both based in Cambridge, UK, and their colleagues, who have taken the position that semantic dementia and the fluent variant of primary progressive aphasia may be “two sides of the same coin.” They have framed the debate as follows. Are there “subtle non-verbal semantic deficits ... in fPPA [(fluent primary progressive aphasia)][that are] fundamental to the anomia and reflect, therefore, damage to an amodal semantic system associated with bilateral, though often asymmetric, anterior temporal lobe (ATL) atrophy?” Alternatively, does semantic dementia “result from a disease process that encompasses two separate neurocognitive networks – a left hemisphere language network and a bilateral fusiform network for face and object recognition – which, though jointly compromised in SD [i.e., “semantic

dementia”], can also be damaged independently?” (Adlam et al., 2006, p. 3067).

Nonverbal deficits in fluent primary progressive aphasia?

To test between the two alternatives laid out in the previous paragraph, Adlam et al. (2006) recruited seven patients who met the diagnostic criteria of early-stage fluent primary progressive aphasia. Along with standard neuropsychological tests of language, memory, and other cognitive functions, they also administered tasks intended to probe their earlier findings that semantic dementia progressively restricts the conceptual knowledge of patients to general and typical features. Examples of these tasks are shown in *Figure 12.5*. The results of the behavioral testing indicated that, on the standardized tests, patients demonstrated a more marked verbal than nonverbal impairment. This is an oft-replicated finding, and follows from the fact that these patients had been selected for meeting diagnostic criteria for fluent primary progressive aphasia. Results from the tests designed to assess the selective loss of specific knowledge, particularly about stimuli with properties that are atypical of their category, on the other hand, contradicted Mesulam’s assertion that “language is the only area of prominent dysfunction for at least the first two years of the disease.” On the Levels of Specificity and Typicality (LOST) word-matching task (*Figure 12.5.A* and *B*), for example, the patients were selectively impaired at matching low-typical targets among close distractors. On the object-color decision task (*Figure 12.5.C*), patients were only impaired on NR > R trials, that is, when the foil was a more domain-typical color than the target. The authors interpreted these findings as evidence that fluent primary progressive aphasia/semantic dementia impairs conceptual knowledge, in addition to language functions, at early stages of the disease:

In summary of this issue: how are we to classify patients like the ones studied here, who present with a prominent fluent, anomic aphasia and a less prominent but still significant agnosia that is modulated by concept familiarity/typicality? These patients qualify for a diagnosis of fPPA on the basis of their preserved everyday function and the predominance of their aphasia; but (i) when

The locus of damage in fluent primary progressive aphasia?

FIGURE 12.5 Testing semantic impairments in fluent primary progressive aphasia/semantic dementia. Source: Adlam, Patterson, Rogers, Nestor, Salmond, Acosta-Cabronero, and Hodges, 2006. Reproduced with permission of Oxford University Press.

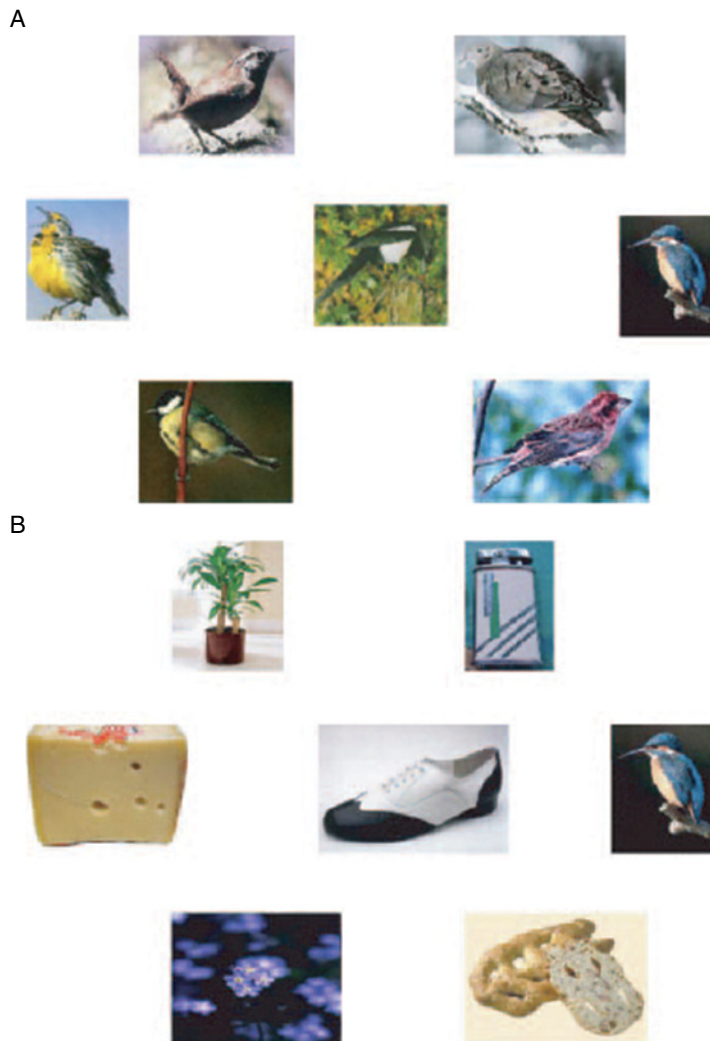


FIGURE 12.5. A AND B Two trials from the Levels of Specificity and Typicality (LOST) word-matching task, illustrating the manipulation of specificity, in that subjects are asked point to the kingfisher (a low-typicality target from the category “birds”) from among a set of close semantic distractors (**A**) and unrelated semantic distractors (**B**). An example of a high-typicality target bird would be a robin.

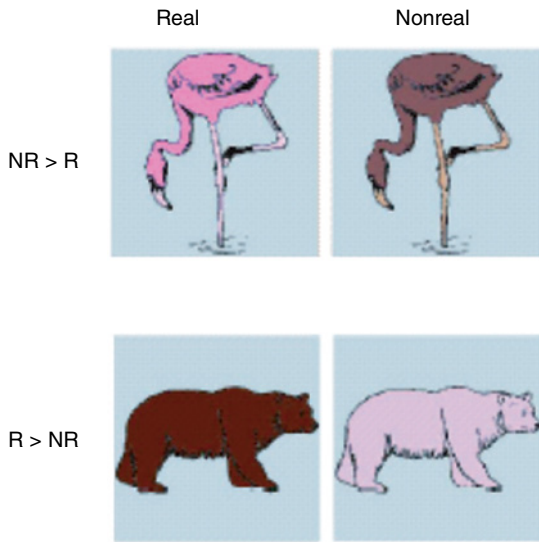


FIGURE 12.5.C Two sets of stimuli from the object-color decision task from Adlam et al. (2006). “NR>R” denotes a trial when the nonreal item has a color that is more typical of the category than the real item (i.e., birds are more typically brown than pink; however, flamingoes are pink), and R>NR a trial when the real item has a color that is more typical of the category than the nonreal item (i.e., mammals are more typically brown than pink; and bears are brown).

neuroimaging studies of patients who they had defined as having semantic dementia. Additionally, a PDP model by Rogers and McClelland (2004) suggests the existence of “a set of shared representation units that tie together all of an object’s properties across different information types.” After damage to or degeneration in these units, “distinctions between items that are very similar semantically tend to be lost ... while distinctions between highly dissimilar objects are maintained” (pp. 378–379). (Note that this is the pattern of deficit observed in the patients of Adlam et al., 2006.) In contrast, most prior research explicitly performed on patients diagnosed with fluent primary progressive aphasia had assumed that the language impairments in these patients resulted from atrophy in the left hemisphere language network, particularly “perisylvian” regions of posterior superior temporal gyrus/inferior parietal lobule and inferior prefrontal cortex of the left hemisphere. Further, Mesulam (2003) had argued that fluent primary

progressive aphasia accompanied by associative agnosia would be expected to exhibit degeneration in inferior occipitotemporal cortex to account for the latter.

Results from the voxel-based morphometry (VBM; *Research Spotlight 11.2*) analysis performed by Adlam et al. (2006), illustrated in *Figure 12.6*, show clear evidence for bilateral degeneration in the anterior temporal lobes of these patients. Further, performance of patients on both verbal and nonverbal tasks was predicted by gray-matter density values in the left inferior temporal region marked with cross hairs in *Figure 12.6.C*, suggesting that the integrity of this region may be crucial for performance on their behavioral tests of semantic memory.

How are we to make sense of the absence of any overlap between the predictions of Mesulam (2003) on the one hand and, on the other, those of Adlam et al. (2006) with respect to which areas are expected to be compromised by fluent primary progressive aphasia? One possible answer comes from an earlier PET study of semantic dementia whose results suggested that activity in regions presumably spared the direct effects of degeneration nonetheless showed abnormal patterns of activity.

Distal effects of neurodegeneration

The study of Mummery et al. (1999) compared patients diagnosed with semantic dementia with age- and education-matched control subjects. Results from a VBM analysis were highly similar to those illustrated in *Figure 12.6*, indicating a degenerative focus in anterior temporal lobe and, importantly, a sparing of the gray-matter integrity of the ventral surface of temporo-occipital cortex. This finding was particularly noteworthy in light of the PET results, which indicated that activity related to word matching and to picture matching was significantly reduced in patients, relative to control subjects, in lateral and inferior temporo-occipital cortex (*Figure 12.7*). This finding suggests a way to reconcile the predictions from Mesulam (2003), on the one hand, and Adlam et al. (2006), on the other: neurodegeneration associated with fluent primary progressive aphasia/semantic dementia may directly impact the integrity of anterior temporal cortex, and this structural degeneration may lead to functional abnormalities in the inferior and lateral temporo-occipital cortex, the normal functioning of which is known to be critical for

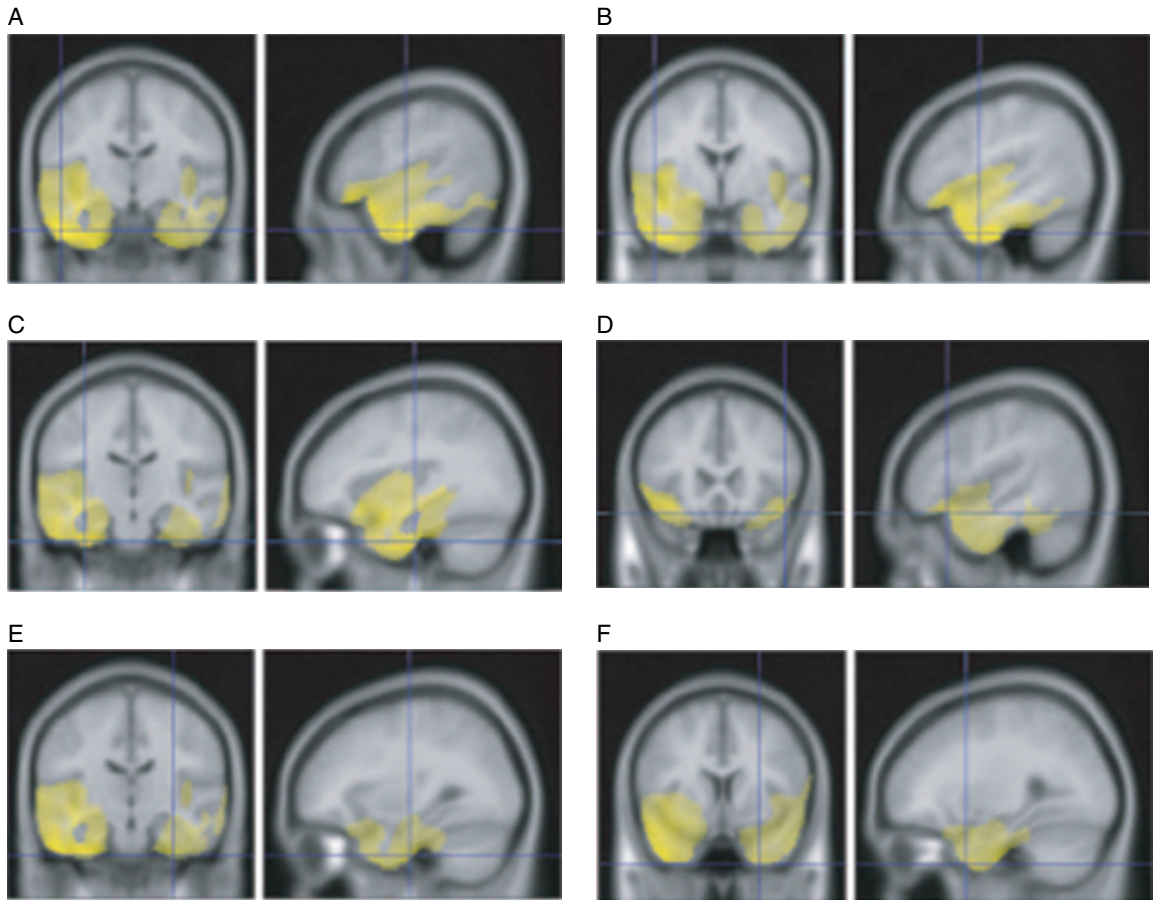


FIGURE 12.6 Results from the whole-brain VBM analysis from Adlam et al. (2006), showing extensive gray matter reduction (relative to control subjects) in each of six patients with fluent primary progressive aphasia. The cross hairs on each image indicate the peak areas of atrophy. Note that these occur in both hemispheres, and exclusively in anterior temporal lobe. Images are shown in neurological display convention (i.e., right side of coronal image corresponds to right hemisphere). Source: Adlam, Patterson, Rogers, Nestor, Salmond, Acosta-Cabronero, and Hodges, 2006. Reproduced with permission of Oxford University Press.

high-level visual perception and, by many accounts, the perceptual features that contribute to the semantic representation of objects.

Systems-level accounts of how disruption of ATL can lead to abnormal functioning in the visual ventral stream include the designation of this region as a critical “hub” for semantic representation (Rogers et al., 2006), and the related idea that it may be a source of “top-down modulatory input necessary for successful retrieval of semantic representations stored in posterior regions” (Martin and Chao, 2001, p. 198). The “hub” model gets further explication in *Web Clip 12.2*.

Entente cordiale

This chapter concludes with an all-too-rare consensus resolution of a hotly debated topic in cognitive neuroscience. An important factor pushing for this outcome was the fact that the stakes in the debate about the neural and cognitive impairments of fluent primary progressive aphasia/semantic dementia have very important implications for the diagnosis and treatment of patients who are suffering from the disease. To achieve it, an international “consensus group” including all of the neurologists whose work we have highlighted here met on three different occasions.

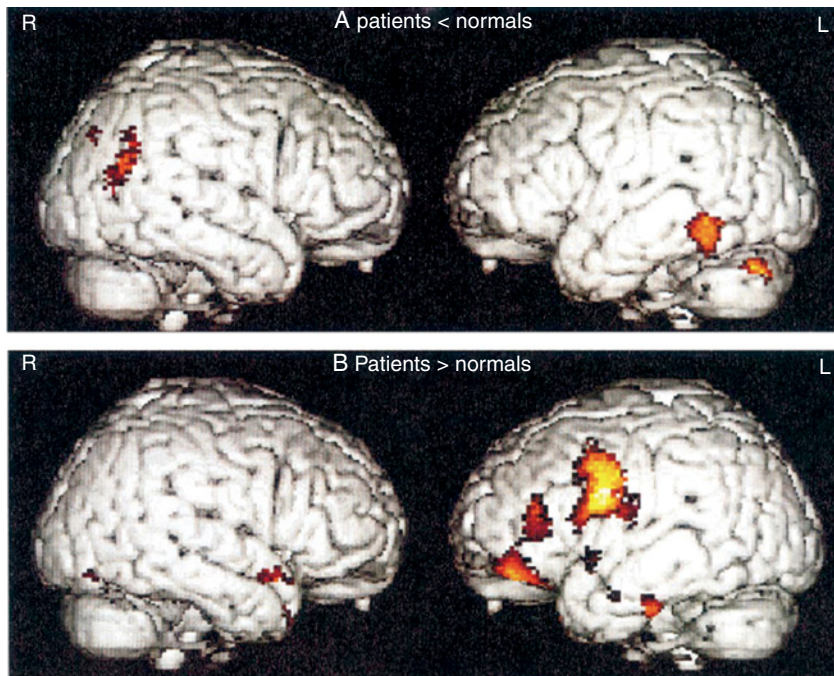


FIGURE 12.7 Difference images from PET scans comparing semantic task-related activity in semantic dementia patients vs. control subjects (“Normals”). Source: Mummery, Patterson, Price, Vandenberg, Price, and Hodges, 1999. Reproduced with permission of Oxford University Press.

Rather than just “getting together to argue,” however, they followed a systematic procedure:

The group reviewed video presentations of 12 PPA cases from different sites. A list of 17 salient speech and language features was provided and each clinician rated whether specific language features were present or not. Videos included a component of spontaneous speech and various portions of formal language evaluations. The analysis of the responses revealed a high level of agreement. There were 15 clinical features in which experts agreed over 80% of the time. Thirteen of these features were included in an operationalized classification scheme listing the main features of the 3 most commonly recognized clinical presentations of PPA. (Gorno-Tempini et al., 2011)

The three subtypes of primary progressive aphasia identified by the consensus group are: nonfluent/agrammatic variant; semantic variant; and logopenic variant. It is on the existence and characteristics of the semantic variant that we have focused in this chapter, and the consensus diagnostic and classification criteria are very similar to what one would predict from the

findings of Mummery et al. (1999) and Adlam et al. (2006). *Web Link 12.1* features such a patient.

NUANCE AND CHALLENGES

I told you it would be like this! As cognitive functions become more abstract, it becomes increasingly difficult to design studies that unequivocally measure the construct of primary interest. The tradeoff, of course, is that what one gives up in reductive clarity, one gains in being able to grapple with questions that lie at the very heart of what it means to be human (from the perspective of “Je pense, donc je suis”). Additionally, although we haven’t given it much explicit consideration here, semantic memory is one of the domains in which cognitive neuroscience makes considerable contact with the cognitive psychology school of embodied cognition, a perspective that is very prominent in the zeitgeist, and that many feel might unify the study of cognition from psychological, cognitive neuroscience, and philosophical traditions.

END-OF-CHAPTER QUESTIONS

1. What are three interpretations of the fact that patients have been characterized with a selective deficit in the ability to name nonliving things and others with a selective deficit in the ability to name living things?
2. Summarize the arguments for and against placing a high level of importance on the findings from a single case study, and on the findings from two case studies that together form a double dissociation?
3. Name at least two possible explanations when a neuroimaging study finds selectively elevated activity for the naming of the color of an object relative to the naming of its shape, or of an action that one could take with that object, or some other property. What additional evidence can be drawn on to constrain the interpretation (e.g., to rule out alternative explanations)?
4. How has fMRI adaptation been used to argue for a functionally different role for left mid-fusiform gyrus relative to its right hemisphere homolog and to areas of lateral occipital cortex?
5. How do neurodegenerative disorders, such as variants of frontotemporal dementia, differ from acute-episode neurological insults, such as stroke or viral encephalitis?
6. What pattern(s) of behavioral performance has (have) been pointed to as evidence for an inherently conceptual, as opposed to lexical, deficit in the semantic variant of primary progressive aphasia?
7. What brain area is selectively compromised in early stages of the semantic variant of primary progressive aphasia? What function, relevant to semantic dementia, is this region believed to support?
8. What are salient differences between the consensus classifications of the nonfluent/agrammatical variant of primary progressive aphasia and the semantic variant of primary progressive aphasia? Despite these differences, how do the two differ in more fundamental ways from Alzheimer's disease?

REFERENCES

- Adlam, Anna-Lynne R., Karalyn Patterson, Timothy T. Rogers, Peter Nestor, Claire H. Salmond, Julio Acosta-Cabrero, and John R. Hodges. 2006. "Semantic Dementia and Fluent Primary Progressive Aphasia: Two Sides of the Same Coin?" *Brain* 129 (11): 3066–3080. doi:10.1093/brain/awl285.
- Caramazza, Alfonso, and Jennifer R. Shelton. 1998. "Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction." *Journal of Cognitive Neuroscience* 10 (1): 1–34. doi:10.1093/neucas/4.4.399-k.
- Damasio, Hanna, Thomas J. Grabowski, Daniel Tranel, Richard D. Hichwa, and Antonio R. Damasio. 1996. "A Neural Basis for Lexical Retrieval." *Nature* 380 (6574): 499–505. doi:10.1038/380499a0.
- Farah, Martha J., and James L. McClelland. 1991. "A Computational Model of Semantic Memory Impairment: Modality Specificity and Emergent Category Specificity." *Journal of Experimental Psychology: General* 120 (4): 339–357.
- Farah, Martha J., Michael M. Meyer, and Patricia A. McMullen. 1996. "The Living/Nonliving Dissociation is Not an Artifact: Giving an A Priori Implausible Hypothesis a Strong Test." *Cognitive Neuropsychology* 13 (1): 137–154. doi:10.1080/026432996382097.
- Gorno-Tempini, Maria L., Argye E. Hillis, Sandra Weintraub, Andrew Kertesz, Mario F. Mendez, Stefano F. Cappa, ... Murray Grossman. 2011. "Classification of Primary Progressive Aphasia and Its Variants." *Neurology* 76 (11): 1006–1014. doi:10.1212/WNL.0b013e31821103e6.
- Grill-Spector, Kalanit, Richard Henson, and Alex Martin. 2006. "Repetition and the Brain: Neural Models of Stimulus-Specific Effects." *Trends in Cognitive Sciences* 10 (1): 14–23. doi:10.1016/j.tics.2005.11.006.
- Mahon, Bradford Z., and Alfonso Caramazza. 2009. "Concepts and Categories: A Cognitive Neuropsychological Perspective." *Annual Review of Psychology* 60: 27–51. doi:10.1146/annurev.psych.60.110707.163532.
- Martin, Alex. 2007. "The Representation of Object Concepts in the Brain." *Annual Review of Psychology* 58: 25–45. doi:10.1146/annurev.psych.57.102904.190143.
- Martin, Alex, and Linda L. Chao. 2001. "Semantic Memory and the Brain: Structure and Processes." *Current Opinion in Neurobiology* 11 (2): 194–201. doi:10.1016/S0959-4388(00)00196-3.

- Martin, Alex, James V. Haxby, François M. Lalonde, Cheri L. Wiggs, and Leslie G. Ungerleider. 1995. "Discrete Cortical Regions Associated with Knowledge of Color and Knowledge of Action." *Science* 270 (5233): 102–105. doi:10.1126/science.270.5233.102.
- Martin, Alex, Leslie G. Ungerleider, and James V. Haxby. 2000. "Category Specificity and the Brain: The Sensory/Motor Model of Semantic Representations of Objects." In *The New Cognitive Neurosciences*, 2nd ed., edited by Michael S. Gazzaniga, 1023–1036. Cambridge, MA: MIT Press.
- Martin, Alex, Cheri L. Wiggs, Leslie G. Ungerleider, and James V. Haxby. 1996. "Neural Correlates of Category-Specific Knowledge." *Nature* 379 (6566): 646–652. doi:10.1038/379649a0.
- McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly. 1995. "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory." *Psychological Review* 102 (3): 419–457. doi:10.1037/0033-295X.102.3.419.
- Mesulam, M.-Marsel. 1982. "Slowly Progressive Aphasia without Generalized Dementia." *Annals of Neurology* 11 (6): 592–598. doi:10.1002/ana.410110607.
- Mesulam, M.-Marsel. 2003. "Primary Progressive Aphasia – A Language-Based Dementia." *New England Journal of Medicine* 346 (16): 1535–1542. doi:10.1056/NEJMr022435.
- Mummery, Catherine J., Karalyn Patterson, Richard J. Price, R. Vandenberg, Cathy J. Price, and John R. Hodges. 1999. "Disrupted Temporal Lobe Connections in Semantic Dementia." *Brain* 122 (1): 61–73. doi:10.1093/brain/122.1.61.
- Petersen, Steven E., Peter T. Fox, Michael I. Posner, Mark Mintun, and Marcus E. Raichle. 1988. "Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing." *Nature* 331 (6157): 585–589. doi:10.1038/331585a0.
- Rogers, Timothy T., Julia Hocking, Uta Noppeney, Andrea Mechelli, Maria Luisa Gorno-Tempini, Karalyn Patterson, and Cathy J. Price. 2006. "Anterior Temporal Cortex and Semantic Memory: Reconciling findings from Neuropsychology and Functional Imaging." *Cognitive, Affective, and Behavioral Neuroscience* 6 (3): 201–213. doi:10.3758/CABN.6.3.201.
- Rogers, Timothy T., and James L. McClelland. 2004. *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.
- Rogers, Timothy T., and David C. Plaut. 2002. "Connectionist Perspectives on Category-Specific Deficits." In *Category Specificity in Brain and Mind*, edited by Emer M. E. Forde and Glyn W. Humphreys, 251–289. Hove, UK: Psychology Press.
- Rosenbaum, R. Shayna, Stefan Köhler, Daniel L. Schacter, Morris Moscovitch, Robyn Westmacott, Sandra E. Black, ... Endel Tulving. 2005. "The Case of K. C.: Contributions of a Memory-Impaired Person to Memory Theory." *Neuropsychologia* 43 (7): 989–1021. doi:10.1016/j.neuropsychologia.2004.10.007.
- Simons, Jon S., Wilma Koutstaal, Steve Prince, Anthony D. Wagner, and Daniel L. Schacter. 2003. "Neural Mechanisms of Visual Object Priming: Evidence for Perceptual and Semantic Distinctions in Fusiform Cortex." *NeuroImage* 19 (3): 613–626. doi:10.1016/S1053-8119(03)00096-X.
- Warrington, Elizabeth K., and Rosaleen McCarthy. 1983. "Category Specific Access Dysphasia." *Brain* 106 (4): 859–878. doi:10.1093/brain/106.4.859.
- Warrington, Elizabeth K., and Rosaleen McCarthy. 1987. "Categories of Knowledge: Further Fractionations and an Attempted Integration." *Brain* 110 (5): 1273–1296. doi:10.1093/brain/110.5.1273.
- Warrington, Elizabeth K., and Tim Shallice. 1984. "Category Specific Semantic Impairments." *Brain* 107 (3): 829–853. doi:10.1093/brain/107.3.829.

OTHER SOURCES USED

- Buckner, Randy L., Julie Goodman, Marc Burock, Michael Rotte, Wilma Koutstaal, Daniel L. Schacter, ... Anders M. Dale. 1998. "Functional-Anatomic Correlates of Object Priming in Humans Revealed by Rapid Presentation Event-Related fMRI." *Neuron* 20 (2): 285–296. doi:10.1016/S0896-6273(00)80456-0.
- Buckner, Randy L., Steven E. Petersen, Jeffrey G. Ojemann, Francis M. Miezin, Larry R. Squire, and Marcus E. Raichle. 1995. "Functional Anatomical Studies of Explicit and Implicit Memory Retrieval Tasks." *Journal of Neuroscience* 15 (1): 12–29.
- Chao, Linda L., James V. Haxby, and Alex Martin. 1999. "Attribute-Based Neural Substrates in Temporal Cortex for Perceiving and Knowing about Objects." *Nature Neuroscience* 2 (10): 913–919. doi:10.1038/13217.
- Kensinger, Elizabeth A., Michael T. Ullman, and Suzanne Corkin. 2001. "Bilateral Medial Temporal Lobe Damage Does Not Affect Lexical or Grammatical Processing: Evidence from Amnesic Patient H. M." *Hippocampus* 11 (4): 347–360. doi:10.1002/hipo.1049.
- Martin, Alex. 2009. "Circuits in Mind: The Neural Foundations for Object Concepts." In *The Cognitive*

- Neurosciences*, 4th ed., edited by Michael S. Gazzaniga, 1031–1045. Cambridge, MA: MIT Press.
- McCarthy, Rosaleen A., and Elizabeth K. Warrington. 1988. "Evidence for Modality-Specific Meaning Systems in the Brain." *Nature* 334 (6181): 428–430. doi:10.1038/334428a0.
- Mummery, Catherine J., Karalyn Patterson, Cathy J. Price, John Ashburner, Richard S. J. Frackowiak, and John R. Hodges. 2000. "A Voxel-Based Morphometry Study of Semantic Dementia: Relationship between Temporal Lobe Atrophy and Semantic Memory." *Annals of Neurology* 47 (1): 36–45. doi:10.1002/1531-8249(200001)47:1 <36::AID-ANA8>3.0.CO;2-L.
- Paulesu, Erardo, John Harrison, Simon Baron-Cohen, J. D. G. Watson, L. Goldstein, J. Heather, ... Chris D. Frith. 1995. "The Physiology of Coloured Hearing: A PET Activation Study of Colour–Word Synaesthesia." *Brain* 118 (3): 661–676. doi:10.1093/brain/118.3.661.
- Plaut, David C. 1995. "Double Dissociation without Modularity: Evidence from Connectionist Neuropsychology." *Journal of Clinical and Experimental Neuropsychology* 17 (2): 291–321. doi:10.1080/01688639508405124.
- Schacter, Daniel L., and Randy L. Buckner. 1998. "Priming and the Brain." *Neuron* 20 (2): 185–195. doi:10.1016/S0896-6273(00)80448-1.
- Schwartz, Myrna F., Oscar S. M. Marin, and Eleanor M. Saffran. 1979. "Dissociations of Language Function in Dementia: A Case Study." *Brain and Language* 7 (3): 277–306. doi:10.1016/0093-934X(79)90024-5.
- Simner, Julia, Catherine Mulvenna, Noam Sagiv, Elias Tsakanikos, Sarah A. Witherby, Christine Fraser, Kirsten Scott, and Jamie Ward. 2006. "Synaesthesia: The Prevalence of Atypical Cross-Modal Experiences." *Perception* 35 (8): 1024–1033. doi:10.1068/p5469.
- Warrington, Elizabeth K. 1975. "Selective Impairment of Semantic Memory." *Quarterly Journal of Experimental Psychology* 27 (4): 635–657. doi: 10.1080/14640747508400525.

FURTHER READING

- Gorno-Tempini, Maria L., Argye E. Hillis, Sandra Weintraub, Andrew Kertesz, Mario F. Mendez, Stefano F. Cappa, ... Murray Grossman. 2011. "Classification of Primary Progressive Aphasia and Its Variants." *Neurology* 76 (11): 1006–1014. doi:10.1212/WNL.0b013e31821103e6.
The consensus paper on PPA and semantic dementia.
- Grill-Spector, Kalanit, Richard Henson, and Alex Martin. 2006. "Repetition and the Brain: Neural Models of Stimulus-Specific Effects." *Trends in Cognitive Sciences* 10 (1): 14–23. doi:10.1016/j.tics.2005.11.006.
No consensus, but nonetheless constructive; three scientists with different explanations for repetition effects seen with fMRI describe their models and articulate a way forward for adjudicating between them.
- Lewis, Penelope A., and Simon J. Durrant. 2011. "Overlapping Memory Replay during Sleep Builds Cognitive Schemata." *Trends in Cognitive Sciences* 15 (8): 343–351. doi:10.1016/j.tics.2011.06.004.
Proposal for how information processing during sleep may contribute to the semanticization of long-term memories.
- Mahon, Bradford Z., and Alfonso Caramazza. 2009. "Concepts and Categories: A Cognitive Neuropsychological Perspective." *Annual Review of Psychology* 60: 27–51. doi:10.1146/annurev.psych.60.110707.163532.
The perspective of two prominent advocates of the category-specific models of semantic memory.
- Martin, Alex. 2007. "The Representation of Object Concepts in the Brain." *Annual Review of Psychology* 58: 25–45. doi:10.1146/annurev.psych.57.102904.190143.
Comprehensive review from one of the leading cognitive neuroscientists of semantic memory.
- Rogers, Timothy T., and James L. McClelland. 2008. "Précis of *Semantic Cognition: A Parallel Distributed Processing Approach*." *Behavioral and Brain Sciences* 31 (6): 689–749. doi:10.1017/S0140525X0800589X.
Précis of Rogers and McClelland's PDP model of semantic cognition, followed by many critiques by others working in the field, followed by a rejoinder from R & M.

CHAPTER 13

SHORT-TERM AND WORKING MEMORY

KEY THEMES

- Short-term memory (STM) refers to the span of information (e.g., a telephone number, an array of objects) that one can hold in mind without assistance from the MTL. Although it is quickly lost when attention is directed elsewhere, it can, in principle, be maintained indefinitely.
- Working memory refers to the mental manipulation of information in STM, such as performing mental arithmetic, or reordering items into numerical or alphabetical order.
- From the 1940s through the 1960s, lesion findings indicated that the PFC was important for performance on working memory tasks, but not for STM per se.
- Despite the previous point, the 1971 extracellular electrophysiology studies reporting elevated delay-period-spanning activity of PFC neurons was interpreted by some (although not the authors themselves) as a neural correlate of STM.
- The conceptual integration of working memory theory from cognitive psychology with neurophysiological findings from PFC led to advances in working memory theory, and to the idea that PFC houses domain-segregated STM buffers.
- Subsequent lesion and electrophysiological studies have led to a reinterpretation of the functions of PFC in terms of working memory, rather than STM.
- The superior sensitivity of MVPA over univariate analysis methods has provided compelling evidence that the representation of information held in STM is supported in regions also responsible for its perception, including V1 for visual STM.
- The superior specificity of MVPA over univariate analysis methods suggests that elevated delay-period activity does not always correspond to stimulus representation per se.
- Retrospective MVPA of extracellular recording data is revealing that PFC follows a dynamic trajectory through high-dimensional state space in support of each phase of working memory task performance.

CONTENTS

KEY THEMES

“PROLONGED PERCEPTION” OR “ACTIVATED LTM?”

DEFINITIONS

ELEVATED, SUSTAINED ACTIVITY

Early focus on role of PFC in the control of STM

Single-unit delay-period activity in PFC and
thalamus

*The implications, and influence, of these first
studies of PFC electrophysiology*

Working memory

*The extracellular electrophysiology of “working
memory”*

*A third(?) generation of PFC lesion studies
STM vs. working memory*

A BRAVE NEW WORLD OF MULTIVARIATE DATA ANALYSIS

The tradition of univariate analyses

MVPA of fMRI

*Increased sensitivity reveals delay-period
storage in “silent” areas*

*Increased specificity reveals absence of
stimulus information in elevated delay-
period activity*

The neural bases of STM

Retrospective MVPA of single-unit extracellular
recordings

*A population-level analysis of “Dynamic
Coding for Cognitive Control in Prefrontal
Cortex”*

THREE-QUARTERS OF A CENTURY

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

“PROLONGED PERCEPTION” OR “ACTIVATED LTM?”

Although the order of chapters in this book by and large follows what one finds in most textbooks on cognitive psychology and cognitive neuroscience, the placement of this one is at variance with its typical placement at the transition point between a just-completed section on sensation/perception/attention and an ensuing section on LTM. This scheme, in effect, portrays STM as “prolonged perception.” My rationale for positioning the chapter that you are currently reading here, instead of there, is twofold. First, it is increasingly the case that short-term and working memory are understood as being supported by the activation of semantic, as well as perceptual, representations. (Additionally, the idea of a role for the hippocampus in some types of short-term memory is gaining prominence in the literature.) Thus, having an understanding of the organization and neural substrates of LTM will be important for what’s to follow in this chapter. Second, short-term and working memory are closely intertwined with cognitive control and other aspects of high-level cognition that will be the focus of the final section of this book, and so this chapter will serve as a transition to this final section.

DEFINITIONS

Already, in the preamble that you just read, you may have noted the awkward formulation of “short-term and working memory.” (And you may have thought *Make up your mind!*) Well, here’s the thing: it’s complicated. Different people use these terms to mean different things, sometimes in different contexts. *Short-term memory* (STM) has a long history going back at least as far as William James and Ebbinghaus. To experimental psychologists, STM is expressed when subjects are asked to recognize or recall a small number of items more-or-less immediately after these items were presented. (Indeed, in some contexts the terms “immediate memory” and STM are synonymous.) Psychologists and neuropsychologists measure

the capacity of verbal STM with variants of the digit span test, in which the subject first hears the digits “5 ... 2” and must repeat them back to the experimenter. Assuming success, she next hears “7 ... 9 ... 4” and repeats these back. This process continues, with the span of digits increasing by one item, and the subject given two attempts at each span, until she makes an error on two consecutive attempts at the same length. The largest number of items that she was able to recall is referred to as her “digit span.” The psychologist George Miller (1920–2012), often credited as one of the progenitors of cognitive neuroscience as a formal discipline, famously noted the high probability that this subject’s span would fall within that of “the magical number 7 \pm 2.”

This seems fairly straightforward, but what if after hearing each series of digits, the subject was asked to wait 10 seconds before reciting them back? What about 10 minutes? 10 hours? At what point would her performance stop reflecting STM? Alternatively, what if she were asked to recite the digits back immediately, but to do so in reverse order (“backward digit span”)? Or to recite them back in their cardinal order, from smallest to largest (i.e., “7 ... 9 ... 4” becomes “4, 7, 9”)? Or to mentally calculate the sum of the remembered digits and respond with that? We can begin to address some of these questions with a vignette featuring this author’s go-to patient for explaining all things mnemonic:

One of the privileges afforded graduate students at MIT’s Behavioral Neuroscience Laboratory has been the opportunity to transport ... H. M. from and to his home, a handful of hours distant, for his roughly semiannual research visit ... The convention for H. M.’s transport was for the designated graduate student to travel with a companion, and there was never a shortage of volunteers (typically a fellow student or a postdoc) eager to take a road trip with this famous patient. On one such trip the two scientists-in-training and their charge sought to pass the time by playing a game in which each player selects a color – on this occasion, green, blue, and white – and accumulates points for each car painted in his or her color that passes in the opposite direction on the highway. Each player counts aloud, and the gaps between passing cars are typically filled with cheering and good-natured banter. H. M. participated fully in the

game, selecting his color, accurately keeping track of his running total, and participating in the debate about whether a teal-colored car should be scored as blue, green, or neither. Indeed, on this occasion H. M. won, accruing a score of 20 first. A round of congratulations was exchanged, followed by a lull as the car rolled through the undulating central Massachusetts countryside. A few minutes later, the guest traveler, eager to maximize his once-in-a-lifetime opportunity to gain first-hand insight from this famous patient, asked “Henry, what are you thinking about right now?” H. M. replied that his count of white cars had now increased to 36.

The driver and guest were both impressed that this patient, famously incapable of remembering virtually anything that had occurred in his life since his 1953 surgery, had accurately maintained and updated a running count of arbitrarily selected “target stimuli” across a span of several minutes, with no evident source of external support or reinforcement. The three travelers commented on this before the guest traveler redirected the conversation to a line of questions that was typical of these trips: “Do you know what today’s date is?” “Do you know who the current President is?” “Do you know who we are; where we’re going today?” H. M. complied with good-natured responses, as always, clearly enjoying the interaction with and attention from these young, engaged travelling companions. Very quickly, however, H. M. initiated another typical element in the driving-with-H.M. script, by steering the conversation towards a reminiscence from his youth, the portion of his life still mentally accessible after his surgery. (The story, about riding the Silver Meteor passenger train on a multi-day trip to visit an aunt in Florida, had already been told several times at that point in the trip, a product of the teller not remembering the previous tellings.) At that point, sensing a “teachable moment,” the driver of the car interjected with a question of his own: “Henry, do you remember what game we were playing a few minutes ago?” No, he didn’t. “It involved counting cars of different colors; do you remember what your color was?” No. “The three colors were green, blue, and white; do you remember which was yours?” No, the cues didn’t help. “Do you remember who won the game?” No recollection even of the triumph that had, only a few minutes before, produced in H. M. a modest chuckle and satisfied smile. (Postle, 2009, pp. 326–327)

The main point here is that the “short” in short-term memory does not relate to the passage of time,

per se. So long as H. M. continued thinking about the car-counting game, he could have maintained a memory for the number 36 indefinitely. And the fact that *continuous thought* is a key factor is illustrated in the fact that just a few moments of thinking about other things removed any trace of memory of even the circumstances that gave rise to the number 36, not to mention the number itself. And so, in a matter of seconds, what had been a STM that was not dependent on the integrity of the medial temporal lobe (MTL) became a piece of information that, in a neurologically healthy individual, would have been retrieved via long-term memory. This means that the informal use of “short-term memory” to refer to memory for things that happened earlier today, or earlier this week, is neurobiologically inaccurate (*Tip 13.1*).

Working memory can be understood as performance on tests of STM that require the mental manipulation of the remembered information, such as the backward digit span, the reorder-into-numerical-order span, and mental arithmetic. Thus, tests of working memory necessarily include a STM component. As we make our way through this chapter, we shall see that this fact has been a source of considerable confusion, and debate, in this field of research.

A final point that can be made here isn’t specific to the study of STM, but nonetheless provides a convenient springboard into the meat of the chapter: one needs to be careful not to conflate performance on a task with the state of a hypothesized cognitive construct. Adapted to be the moral of *Thought Box 13.1*, one might rephrase this as: *a deficit on an STM task need not imply an STM deficit*.

Tip 13.1

It’s okay, and understandable, when friends or TV newscasters misuse “STM,” but, maddeningly, many neurologists and psychiatrists are also guilty of using it in this imprecise way. Confusingly, many neurobiologists studying synaptic plasticity also use “STM” to refer to the early, pre-protein-synthesis-dependent component of LTP.

THOUGHT BOX

13.1 STM and the PFC? The roots of a long and fraught association

By the 1930s, the functions of most regions of cerebral cortex had been determined, at least to a first order of approximation, with the scientific methods that we introduced way back in *Chapter 1*. An exception to this was the prefrontal cortex, the region that, in the primate, is rostral to those in which damage produced paralysis, or other disorders of motor control. To investigate its function, Carlyle Jacobsen, at Yale University, trained two monkeys to perform a delayed-response task in an apparatus quite similar to the WGTA (Jacobsen, 1936; *Figure 5.2*). After watching while one of two covered food wells was baited, the monkey waited for several seconds during which the lowered screen blocked it from seeing or reaching the wells. When the screen was raised, the animal was given one reach with which to displace the cover and, on a correct response, retrieve the food. Pre-lesion, the animals learned to perform the task almost perfectly. After recovery from bilateral surgical removal of the frontal cortex anterior to the arcuate sulcus (i.e., MASSIVE PFC lesions), the animals' performance never deviated from chance (50% correct). That is, they were reduced to guessing. Importantly, Jacobsen included trials on which no delay was imposed between well-baiting and food retrieval. On these, performance improved dramatically, meaning that poor performance on delay trials could be ascribed to neither visual nor motoric nor motivational problems. Thus, a reasonable description of the findings might seem to be that, on delay trials, monkeys with bilateral PFC lesions could not remember which well had been baited. Jacobsen concluded that the prefrontal cortex is responsible for "immediate memory."

The idea that Jacobsen had *localized* immediate memory to the PFC would soon be called into question, however, by a subsequent study performed at the same laboratory. Robert Malmø (1942) replicated the basic procedure from Jacobsen's experiment, but added the additional experimental factor of turning off the lights in the lab on one half of the trials. (The laboratory had remained illuminated throughout Jacobsen's experiment.) Remarkably, this simple manipulation had the effect of "rescuing" the performance of the PFC-lesioned monkeys, in that they performed correctly on roughly 85% of lights-off trials, despite still getting only 50% correct on lights-on trials.

Thus, even though the delayed-response task is a test of STM, it was evidently incorrect to assume that STM, *per se*, had been disrupted by damage to the PFC. (If it had been, then turning off the lights wouldn't have restored performance.) Malmø attributed his findings to an increased susceptibility to interference in the monkeys after bilateral PFC removal. That is, whereas the ability to retain information over a short period of time remained intact, the ability to guide behavior with this information was compromised when other stimuli competed for attention. Twenty years later, in the chapter that introduced the concept, and the term, "working memory" to neuroscience, Pribram, Ahumada, Hartog, and Roos (1964) reinforced this view, asserting that "aspects other than the trace of memory [i.e., STM] were involved by the frontal procedure [i.e., were compromised by PFC lesions]: action at the time of stimulus presentation and distractability were found to be important" (p. 29).

ELEVATED, SUSTAINED ACTIVITY

Early focus on role of PFC in the control of STM

In the wake of Malmö's study (*Thought Box 13.1*), the majority of research on behavior mediated by the PFC concentrated on the control of STM, and the control of behavior that is guided by STM. The latter was shown to be compromised when monkeys with PFC lesions were shown to be impaired on tests of delayed alternation. Delayed alternation differs from delayed response in that no to-be-remembered information is presented at the beginning of the trial. Instead, what the animal must remember is what its response was on the previous trial, whether or not that response was correct (i.e., was rewarded), and what the rules of the task are. On the first trial, the animal sees two covered food wells, and reaches to one of them (we'll call it A). Unbeknownst to it, both food wells are empty on this trial (thereby guaranteeing that the choice on that trial will be incorrect). During the ensuing delay (during which the barrier is lowered) the well that the animal has just selected is baited. This is alternation because, from the animal's point of view, it's as though the food was in well B on the first trial, but it selected A. Now, on the second trial, the food will be in A. In the third trial, it will be in B, and so on. Once the task is underway, the animal needs not only to retain a STM of where the food had been on the previous trial, it needs to overcome the natural tendency to return to that place, a level of control of behavior that we will consider in *Chapter 14*.

In the human, in 1962, Lila Ghent, Mort Mishkin, and Hans-Lukas Teuber found no impairment of forward digit span performance in a group of 24 patients with PFC lesions, although they did find an impairment in a group of 20 patients with lesions in left temporoparietal cortex (more on this in *Chapter 18*). Two years later, another study revealed echoes in the human of what *Thought Box 13.1* had highlighted in the monkey. In it, Milner (1964) reported on the performance of PFC-lesioned patients on tests of 60 second delayed recognition. When tested with novel nonsense shapes, with no reuse of stimuli across trials (i.e., an "open set"), performance of these patients was intact. With other types of stimuli, however, the

patients were tested on closed sets, and with these their performance was impaired. "Closed set" refers to a stimulus pool made up of a finite number of items, such that they are reused over the course of testing. (For example, the digit span test uses a closed set of the digits 0–9; because the test begins with a span of two, and each length is tested twice, the first four-item list in the series is necessarily comprised of items that have already appeared on earlier trials of the test.) From experimental psychology, we know that the use of closed sets increases the influence of proactive interference, a phenomenon whereby retrieval from memory "gets interfered with" by the residual memory traces of no-longer-relevant items from previous trials. On tests of recognition, proactive interference (often abbreviated PI) can produce false-alarm responses, such that the subject incorrectly endorses a memory probe that had appeared on a previous trial, but not the current one. On tests of recall, it can produce intrusions, such that an item from a previous trial is recalled as having appeared on the current one. Because patients with PFC lesions were impaired when tested with closed sets, but not when tested with an open set, Milner (1964) interpreted their impairment on the former as attributable to the control of PI, rather than to STM per se.

Thus, in the 1960s, the role of the PFC in STM was generally understood to be one of the control of the contents of STM, and/or the control of behavior that must be guided by the contents of STM, rather than of the short-term storage, per se, of information. "This all changed" (well, the thinking of many researchers changed) when extracellular electrophysiological techniques were brought to the study of delay-task performance.

Single-unit delay-period activity in PFC and thalamus

In 1971, the groups of Joaquín Fuster and Garrett Alexander at UCLA, and of Kisou Kubota and Hiroaki Niki at Kyoto University (*Tip 13.2*), published the results of recordings from neurons in the PFC while monkeys performed delayed recognition and delayed alternation, respectively. The cortical recording sites for both groups were the banks of the caudal half of the principal sulcus (the area labeled

Tip 13.2

In the process of performing their seminal research on the neurophysiology of high-level cognition at the Primate Research Institute of Kyoto University, Niki and Kubota also trained an entire cohort of Japanese neurophysiologists, many of whom we will encounter here or in subsequent chapters: Watanabe, Funahashi, Komatsu, Sawaguchi, Matsumura, Nakamura, Sakagami, Matsumoto, ...

Tip 13.3

Recall that an analogous phenomenon in the fMRI probe-evoked response of some regions of human PFC was used to illustrate a failure of the assumption of pure insertion (*Figure 8.5.C*)

“PF” in *Figure 8.9*; many studies subsequent to those of Jacobsen and Malmo had indicated that damage restricted to just this region was sufficient to impair delay-task performance). Additionally, Fuster and Alexander (1971) also recorded in the mediodorsal (MD) nucleus of the thalamus, the thalamic station of the lateral frontal cortical-basal ganglia-thalamic circuits illustrated in *Figure 7.3*.

During delayed-response performance, Fuster and Alexander found that many neurons in both PFC and MD displayed elevated firing rates that spanned the duration of the delay period, which varied in length, unpredictably, within a range of 15 to 65 seconds (*Figure 13.1.A*). During delayed-alternation performance, Kubota and Niki (1971) observed two classes of task-related activity: neurons with elevated activity during the delay (“D”); and neurons that became active just prior to, and during, the response period (“E”; presumably because their activity predicted the onset of

activity in the electromyogram (EMG) that was recording from muscles in the animal’s arm?). Interestingly, the response profile of many E cells changed quantitatively when the animals performed a simple alternation task with no delay period interposed between responses – pre-response bursts were of lower intensity (i.e., slower firing rate), and they preceded the motor response by a smaller period of time (*Tip 13.3*).

Both sets of authors provided interpretations of their data that, at the time of the writing of this book, fit well with current thinking about the PFC. Fuster and Alexander (1971):

The temporal pattern of firing frequency observed in prefrontal and thalamic units during cue and delay periods suggest the participation of these units in the acquisition and temporary storage of sensory information which are implicated in delay response performance. Their function, however, does not seem to be the neural coding of information contained in the test cues, at least according to a frequency code, for we have not found any unit showing differential reactions to the two positions of the reward.

It is during the transition from cue to delay that apparently the greatest number of prefrontal units discharge at

FIGURE 13.1 PFC activity during manual delayed-response and manual delayed-alternation performance.

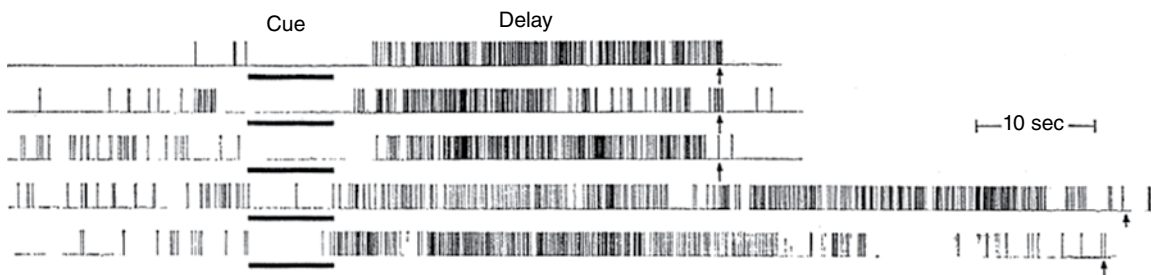


FIGURE 13.1.A Spiking data from five delayed-response trials with delay lengths of 16, 16, 65, and 63 sec., respectively. Source: Fuster and Alexander, 1971. Reproduced with permission of AAAS.

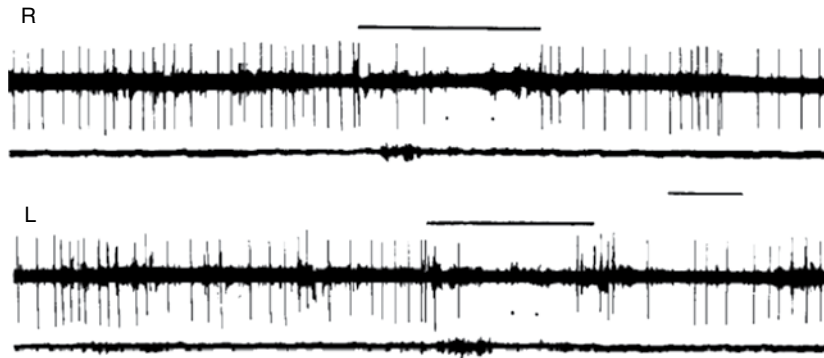


FIGURE 13.1.B Extracellular recording data from a “D” neuron in PFC during periods of delayed-alternation performance leading up to and following a right-hand (R) and a left-hand (L) response. Horizontal bars above each trace indicate time when screen was raised, thereby permitting a response. Below each neuronal trace is the EMG from the left triceps muscle. Dots between the two traces indicate period of juice delivery. Shorter horizontal bar between two sets of traces is a scale bar denoting 0.5 s (Note that *Figure 13.1.B* shows traces from the recoding electrode, with action potentials visible as transient bimodal deviations in the signal, whereas *Figure 13.1.A* shows binarized records of spiking (called “rasters”) that were derived from traces such as those shown in *Figure 13.1.B*.) Source: Kubota and Niki, 1971. Reproduced with permission of The American Physiological Society (APS).

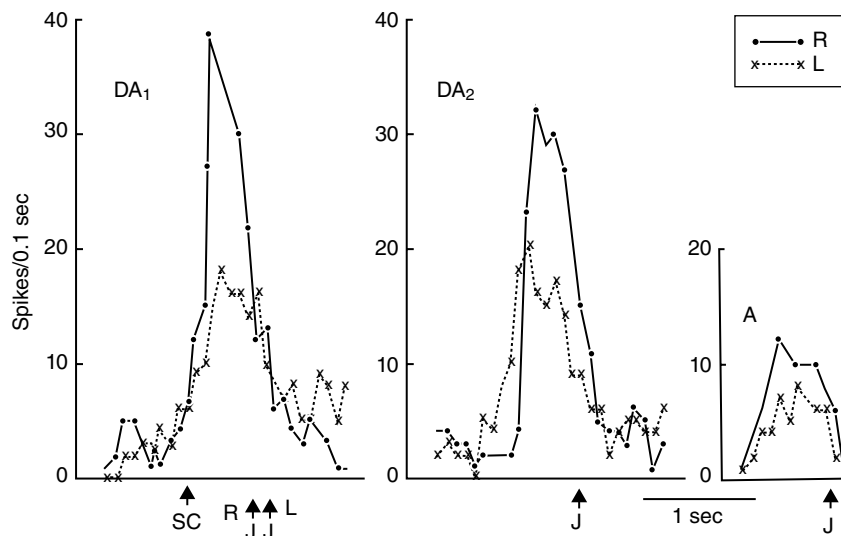


FIGURE 13.1.C Responses of an “E” neuron in PFC during delayed-alternation performance, time-locked to onset of screen raising (SC; plot DA_1), to juice delivery (J; plot DA_2), and to J during simple alternation (i.e., no delays) performance. Source: Kubota and Niki, 1971. Reproduced with permission of The American Physiological Society (APS).

firing levels higher than the intertrial baseline. This may be the basis of the d-c negative potential shift that has been reported to occur at that time in the surface of the prefrontal cortex ... We believe that the excitatory reactions of neurons in MD and granular frontal cortex during delayed response trials are specifically related to the focusing of attention by the animal on information that is being or has been placed in temporary memory storage for prospective utilization. (p. 654)

Kubota and Niki (1971):

in the delay task, periprincipal E units may be causally coupled with the initiation of the voluntary lever pressing. During delay task the E unit activity may be coupled with the initiation and sustaining of ... lever pressing rather than [a] memory retrieval process.

... D units are hardly correlated with the memory storage (1) or retrieval of the memory for the lever pressing (2). Frequency of D unit during delay phases is not apparently different between right and left lever pressings. However, this interpretation does not exclude the possibility that the activities of neurons in the prefrontal cortex represent a memory function related to the choice of the correct performance on the basis of immediate past experience, i.e., remembering the spatially directed response on the preceding trial. (p. 346)

The implications, and influence, of these first studies of PFC electrophysiology

Perhaps to a greater extent than the details of the findings, the very fact of the publication of these studies was compelling, evocative, and highly influential. At the most general level, they represented the first direct studies of brain activity underlying a behavior that was unequivocally *cognitive*. For over a century, of course, scientists had been systematically studying the effect of lesions on various aspects of cognition, but there's somehow something more compelling, more seductive, about "watching the workings of the mind" as compared to inferring them from patterns of behavioral impairment. It was also true that electrophysiologists had been studying brain function for decades. But, prior to 1971, even the most ground-shattering, Nobel Prize-winning of studies were being performed on the "machine-like" functions of perception and motor control – often the animals under investigation were anaesthetized and, even

when not, the reports of these studies contained little to no consideration of the implications of the research for *thinking*. The two papers in question, on the other hand, didn't have to explicitly invoke "thinking"; thinking was, quite simply, what they were about.

With regard to STM, what seems to have made these studies so compelling was, paradoxically, an interpretation that the authors themselves explicitly discounted: some saw in them a physiological correlate of the first of two memory mechanisms that Hebb had postulated in 1949. His theory was, in fact, a dual-trace theory with the first being a "reverberatory" mechanism for "a transient 'memory' of [a] stimulus" (p. 61) that would sustain a representation until it could be encoded into a more permanent state via synaptic strengthening (*Tip 13.4*).

The attribution of explicitly mnemonic interpretations to the 1971 findings became more prevalent when, a few years later, a highly influential model of working memory was introduced in the cognitive psychology literature.

Working memory

The model of working memory proposed in 1974 by the British experimental psychologists Alan Baddeley and Graham Hitch was a response to two factors. First, in their view, contemporary models of STM didn't capture the fact that mental operations performed on information in conscious awareness can be carried out independent of interaction with, or influence on, LTM. (One implication from levels-of-processing theory (*Chapter 11*), for example, was that indefinite "maintenance rehearsal" of information in STM did not predict encoding into LTM. Contemporary models, in the view of Baddeley and Hitch (1974), could not accommodate this fact.)

Tip 13.4

Even though the details of the double helix and protein synthesis were still years in the future, Hebb had intuited that synaptic strengthening couldn't happen instantaneously, without some metabolic events taking place.

Second, they had observed that performance on each of two tasks under dual-task conditions could approach levels of performance under single-task conditions if the two engaged different domains of information, specifically verbal and visuospatial. Thus, early versions of their model called for two STM buffers (dubbed the “phonological loop” and the “visuospatial sketchpad,” respectively) that could operate independently of each other and independently of LTM, although both under control of a Central Executive. This model has proven to be enormously influential, spawning a prodigious amount of research that continues through to this day.

The extracellular electrophysiology of “working memory”

The next major development in the cognitive neuroscience of STM (and working memory) was the proposal, by neuroscientist Patricia Goldman-Rakic (1937–2003), that the sustained delay-period activity in the PFC of the monkey (reported by, e.g., Fuster and Alexander (1971) and Kubota and Niki (1971)) and the storage buffers of the multiple-component model of working memory from cognitive psychology were cross-species manifestations of the same fundamental mental phenomenon. This idea, and the research program that it spawned, are explored in *Research Spotlight 13.1* (Tip 13.5).

RESEARCH SPOTLIGHT

13.1 What vs. where in visual working memory

Earlier in her career, Patricia Goldman-Rakic had trained and collaborated with such luminaries as neuropsychologists Haldor Rosvold and Mortimer Mishkin, and neuroanatomist Walle Nauta. When she moved to Yale University in 1979, she was a leading authority on the anatomy, pharmacology, and functional anatomy (i.e., the effects of lesions) of the PFC. Motivated by the ongoing neurophysiological research coming out of the UCLA and Kyoto groups, on the one hand, and the growing influence of the cognitive model of Baddeley and Hitch on the other, she recruited a former student of Kubota, Shintaro Funahashi, to help her establish an electrophysiology laboratory (Tip 13.2).

The task that Funahashi, Goldman-Rakic, and colleagues used in many of their studies was the oculomotor delayed-response (ODR) task (as introduced and illustrated in *Figure 8.4.*). *Figure 13.2.A* illustrates activity from a single left-hemisphere PFC neuron, recorded during several trials of ODR to each of eight locations. Note that this neuron responds selectively during trials for which a saccade to the downward (270°) target will be required. (Note that a suggestion that it may be also conveying information during trials to other locations comes from the fact that its delay-period activity is suppressed below baseline levels on trials requiring saccades to the three upper visual field locations. One may ask why Funahashi and colleagues found such selectivity in PFC neurons, when neither Fuster and Alexander (1971) nor Kubota and Niki (1971) did. This may have been due, in part, to changes in experimental procedures, including the finer grain of spatial discrimination measured by, and therefore required by, the ODR task.

One important contribution of Goldman-Rakic's theorizing was to suggest that the organization of the visual system into two parallel pathways (recall, *what* vs. *where* had been introduced in the early 1980s) may extend to visual working memory, such that working memory for object identity may be dissociable from visual working memory for object locations. This idea was confirmed in subsequent experimental psychology, neuroimaging, and neuropsychological studies, and led to a “fractionation” of the visuospatial sketchpad, by Baddeley and colleagues, into a “visual cache” (representing *what*) and an “inner scribe” (representing *where/how*) – a clear instance of neuroscience influencing cognitive theory.

Research Spotlight 13.1 *Continued*

In terms of brain organization, Goldman-Rakic's proposal is captured in *Figure 13.2.B*. There are three important points to make about this diagram. First, the heavy arrows representing the flow of information, although schematized cartoons, do reflect the anatomical bias of fibers projecting from posterior regions into PFC, many of these pathways discovered and characterized by Goldman-Rakic and colleagues. (We should note, however, the inaccuracy of suggesting unidirectional patterns of connectivity.) Second, the paper presenting this diagram did also report data from neurons in the inferior convexity of the PFC that showed elevated activity that was preferential for colors or faces over locations in space. Third – and this may well be the only *Research Spotlight* in the entire book for which this happens – most researchers today wouldn't endorse the main point of the model illustrated in *Figure 13.2.B*. Reasons are to follow, but this has simply been so highly influential a model in the cognitive neuroscience of working memory that it couldn't not be included here.

FIGURE 13.2 Studying the organization of visual working memory with ODR tasks.

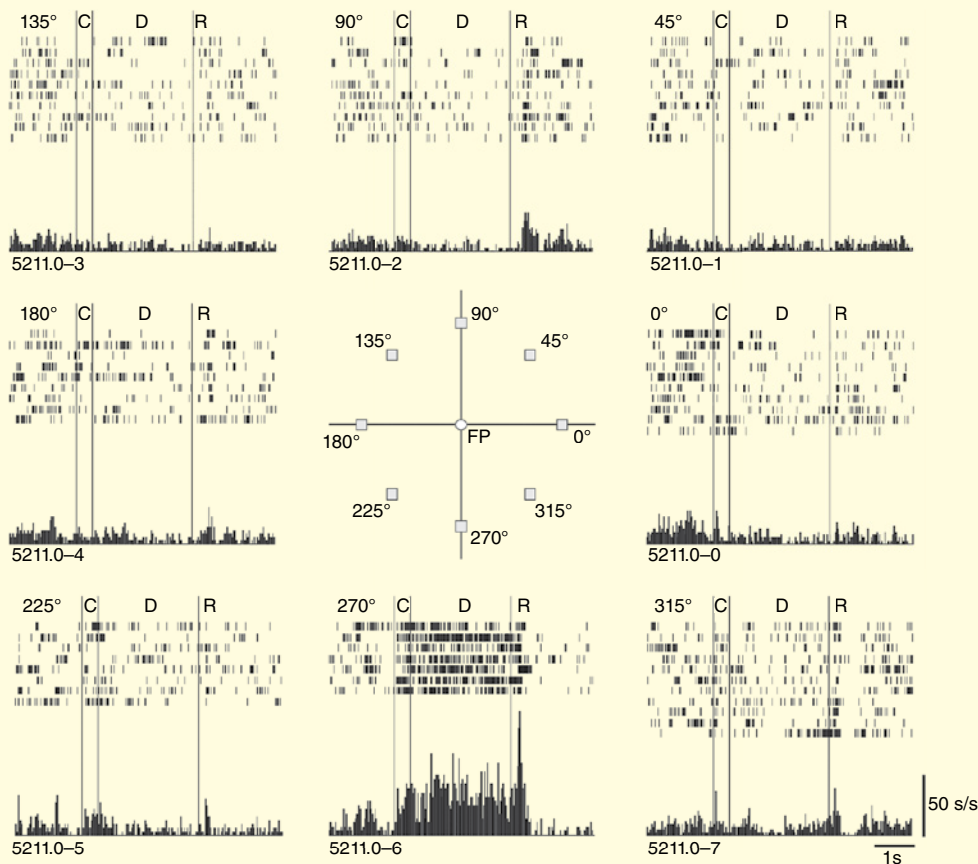


FIGURE 13.2.A The center illustrates the fixation point (FP) and the location of the 8 targets, each 13° of visual angle from FP (and only visible to the animal when flashed briefly to cue a location as the target for an ODR). The plots illustrate firing rate (rows of rasters) for several trials at each location, with a histogram below summing up the activity. Source: Funahashi, Bruce, and Goldman-Rakic. 1989. Reproduced with permission of The American Physiological Society (APS).

Research Spotlight 13.1 Continued

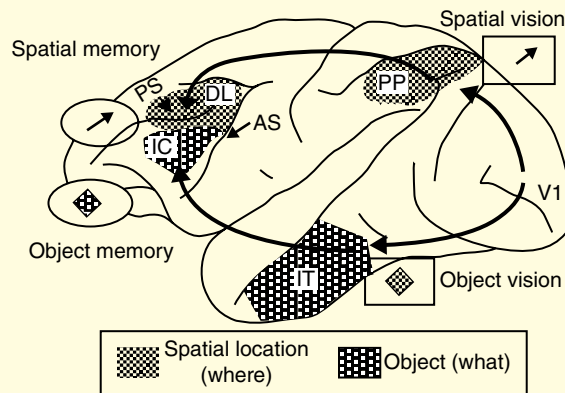


FIGURE 13.2.B Diagram, assembled from the results of many studies from Goldman-Rakic and colleagues, proposing how visual working memory may be supported by extensions into PFC of the where and what pathways. In frontal cortex: PS=principal sulcus; DL=dorsolateral; AS=arcuate sulcus; IC=inferior convexity (the gyrus ventral to the PS). Source: Wilson, Ó Scalaidhe, and Goldman-Rakic 1993. Reproduced with permission of AAAS.

Tip 13.5

Interestingly, the term “working memory” had been proposed as an analogy for PFC function in the early 1960s, in reference to the “noticing order” (a proto version of RAM) in artificial intelligence/computer science research (Pribram et al., 1964; see *Thought Box 13.1*). Perhaps it didn’t catch on in neuroscience research at that time because the technology wasn’t yet available to study it?

As indicated in *Research Spotlight 13.1*, the insight that visual working memory is organized along the lines of the functional organization of the visual system was an important conceptual advance, and led to, among other things, study of neuronal activity in IT cortex and in PPC during delay-task performance. This research, much of it pursued by visual neurophysiologists whose work was considered in *Chapter 5* and *Chapter 6*, followed the precedent established by

Funahashi and colleagues (*Figure 13.2.A*) of proceeding in two stages: first, determine the tuning properties of a neuron; second, study its delay-period activity during trials when the animal is remembering that neuron’s preferred stimulus vs. when it is remembering a non-preferred stimulus.

One set of findings, which relates to the previous chapter’s consideration of repetition effects, came from the laboratory of Robert Desimone, who we first encountered in the context of face processing in IT cortex. Desimone’s then-postdoc, Earl Miller (himself also a former PhD advisee of Charlie Gross), set out to study “how IT neurons retain memories while new information is streaming into the visual system” (Miller, Li, and Desimone, 1991, p. 1377). To study this they used an “ABA” design, in which the animal’s task was to detect the repeat appearance of the first item in a series (designated “A”—there were “ABA,” “ABCA,” and “ABCDA” trials; see trial labeled “Standard” in *Figure 13.3.A*). The principal finding was that neurons tuned for the A stimuli showed decreased stimulus-evoked responses upon the second presentation of A. (Some of these neurons did, and some did not, show

FIGURE 13.3 Stimulus repetition and target match effects in IT and PFC. Source: Miller, Erickson, and Desimone, 1996. Reproduced with permission of the Society of Neuroscience.

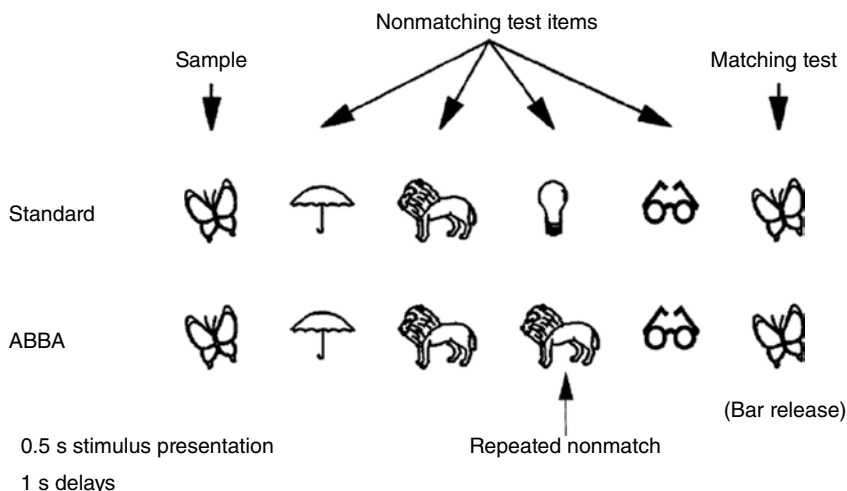


FIGURE 13.3.A Diagrams of the two tasks. The task labeled “Standard” corresponds to the one referred to as “ABA” in text. Trials varied in length with anywhere from 0 to 6 intervening stimuli. Actual stimuli were digitized color images of common objects.

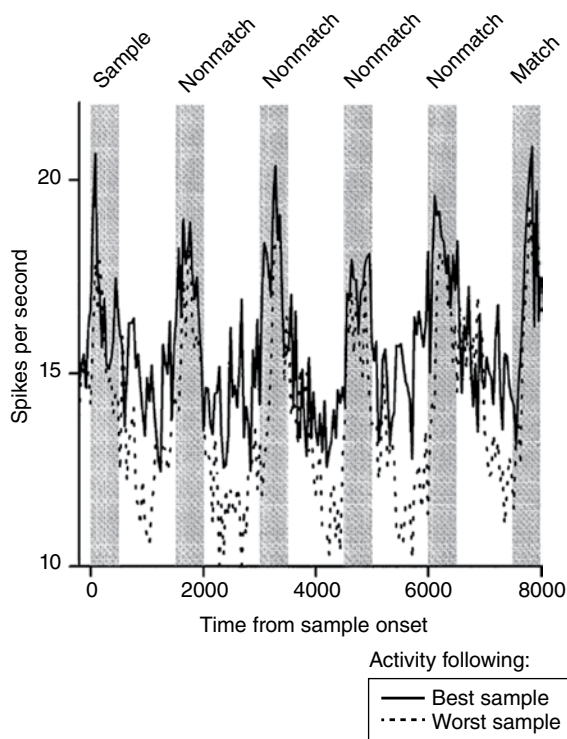


FIGURE 13.3.B Average response (firing rate) histogram from 40 PFC neurons that showed sustained, elevated delay-period activity. The solid trace corresponds to the image that was most effective at driving each neuron (“Best”), the dotted trace to the image least effective at driving it (“Worst”). Note that the vertical axis starts at 10 Hz, the average baseline level of activity for these neurons.

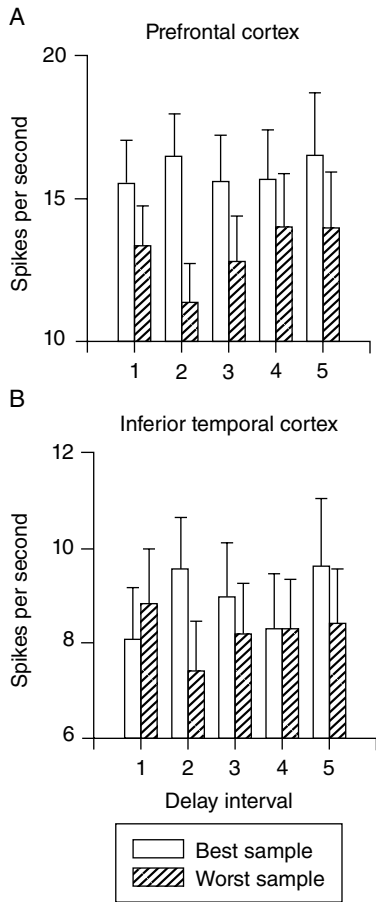


FIGURE 13.3.C Comparison of delay-period activity on trials when the sample was the Best vs. the Worst for neurons in PFC vs. IT.

elevated activity between stimuli.) Based on this, the authors concluded that “the responses of IT neurons could underlie the ability to perform working memory tasks spanning multiple stimulus items” (p. 1378).

Subsequently, however, Miller and Desimone had to revisit this conclusion, after including repetition among irrelevant items (e.g., ABBCA), and finding that monkeys performing near the ceiling with ABA trials performed miserably on what they called “ABBA” trials (*Figure 13.3.A*), invariably responding to the second appearance of the B stimulus as though it were the to-be-remembered item. Thus, the previous supposition about IT neurons “under[lying] the ability to perform working memory tasks” had been

premature, because the monkeys hadn’t been treating ABA trials as working memory trials – holding A in mind in the face of intervening interference until it reappeared. Instead, they had “merely” been detecting repetition in any stimulus. Because the investigators only ever repeated the A stimulus, they weren’t able to see that this was how the monkeys had been performing the task. Further confirmation of this came after the animals learned the ABBA rule with a few more weeks of training. (I.e., it wasn’t that the monkeys couldn’t perform a task requiring detection of the repetition of a particular item, it was just that up until then they hadn’t any reason to do so.) Now, Miller and Desimone (1994) found opposing repetition effects, depending on the behavioral relevance of the repeated stimulus. The repetition of an irrelevant stimulus (“B” in our examples) continued to be suppressed relative to its initial presentation. The repetition of the target stimulus (“A”), in contrast, revealed two classes of neuron: some still produced suppression; but some showed an enhanced response. It would be such “match enhancement” neurons, then, that could underlie the ability to respond correctly on working memory tasks (*Tip 13.6*). But did IT cortex, across the duration of the trial, store the information that was expressed in an IT neuron’s match-enhancement response? This seemed unlikely to Miller and Desimone, because many of the IT neurons from which they had recorded did not show sustained, elevated activity.

For the final study in this series, Miller, Erickson, and Desimone (1996) recorded from the inferior

Tip 13.6

Note that the distinction between the “passive” recognition of the repetition of any stimulus vs. the “active” recognition of a task-relevant target is reminiscent of the constructs of familiarity vs. recollection from the LTM literature (*Chapter 11*). Indeed, some argue that there is no fundamental difference between STM and LTM, in that the same processes may be engaged for retrieving two-second-old memories as well as for two-year-old ones.

convexity of PFC. Figure 13.3.B is a population histogram showing activity of neurons in PFC that did, indeed, show sustained, elevated activity across the entirety of the trial. Figure 13.3.C illustrates that this property was not seen reliably in IT. From these results, Miller, Erickson, and Desimone (1996) noted that the match enhancement effect in IT was effectively a biased competition effect (*Chapter 6*), and they further suggested that the sustained delay-period activity in PFC may be the source of the top-down biasing signal. Further, they noted that “In PF cortex, an explicit neural representation of the sample stimulus appears to be maintained throughout the trial, whereas in IT cortex, it is not. Thus, PF cortex has explicit neural signals correlated with two critical aspects of the DMS task, namely the maintenance of the sample memory trace and an evaluation of whether the test stimulus is a match to it” (p. 5165).

A third(?) generation of PFC lesion studies

In retrospect, there was a fundamental conceptual problem with the research program that led to the idea of working memory buffers localized to the PFC, as depicted in Figure 13.2.B, and as articulated in the previous paragraph. It was hinted at in the title of the previous subsection, by putting the label “working memory” in scare quotes. Many of the tasks used to study mnemonic activity in the PFC (e.g., Figure 13.2.A) have, in fact, been tests of STM – they have required the retention of information across a brief delay period, but no “working” on, or with, that information. Thus, at the level of nomenclature, it was inaccurate to label the activity measured during these tasks as “working-memory activity.” And, as can often happen in science, this lack of terminological precision led to (accompanied, in any event) a conceptual error: the erroneous attribution of the cognitive construct of the STM buffer from the Baddeley and Hitch model to a region (the PFC) whose integrity had been known for decades *not* to be necessary for STM performance (Tip 13.7).

If we consider the studies from the 1940s through to the 1960s as the second generation of PFC lesion studies (the first being those of Ferrier and contemporaries in the second half of the nineteenth century), what might then be considered a third generation of PFC lesion studies was performed in the 1990s in explicit reaction to the model proposed in the Wilson,

Tip 13.7

This is a variant of the “construct validity problem” that is unique to cognitive neuroscience: the localization of a cognitive construct (which, in-and-of-itself, may have high validity) to the wrong part of the brain. Once this occurs, activity from that brain area can be erroneously interpreted as relating to this construct.

Ó Scalaidhe, and Goldman-Rakic paper (1993; Figure 13.2.B). In one, Oxford-based neuroscientists Matthew Rushworth, Richard Passingham, and colleagues demonstrated that STM for object identity is not impaired by bilateral lesions to the inferior convexity of the PFC. In a second, Cypriot-born, Montreal Neurological Institute (MNI)-based neuroscientist Michael Petrides demonstrated a double dissociation whereby principal sulcus lesions did not impair STM for a single stimulus across long (90 and 120 second) delay periods, but did disrupt working memory on a task of 10 second delayed non-repetition of selection from among three, four, or five items. (Called the Self-Ordered Choosing task, the subject first chooses a stimulus (let’s say A) from the pool of [A, B, C], followed by a 10 second delay, followed by a re-presentation of [A, B, C], but now only B or C would be candidate correct choices.) Lesions of anterior IT cortex had the converse effect: impairing STM over long delays but leaving Self-Ordered Choosing performance intact.

How is one to reconcile the seemingly incompatible findings between the electrophysiological findings depicted in Figure 13.2 and Figure 13.3 and the results of these “third generation” lesion studies? Starting from the position of strongest inference, we have to accept that the lesion studies make it clear (again), that performance on simple STM tasks is by and large insensitive to the integrity of the PFC (Tip 13.8). From this, one or two conjectures can follow: (1) PFC is not the only region that supports “an explicit neural representation of the sample stimulus” during the delay period; and/or (2) an “explicit” neural representation of the sample is not needed to succeed on a test of STM. Consistent with (1), stimulus-tuned, distraction-spanning activity has been described in

Tip 13.8

This can be seen as a cautionary tale against being seduced by findings from the newest, most technologically advanced methods (in this instance, single-unit electrophysiology in awake behaving animals) when they contradict facts grounded on an inferentially stronger footing.

anterior IT cortex and in entorhinal cortex. Further, on tests of STM that do not include intervening, distracting stimuli, there's good evidence that sustained stimulus representation may be supported in the sensory networks that processed the sample information. This has been shown electrophysiologically in S1 and in V1, and later in this chapter we will consider fMRI evidence for delay-period stimulus representation in V1. And so PFC delay-period activity, if it does correspond to a mnemonic stimulus representation (to be taken up in *Thought Box 13.2*), may reflect just a part of a broadly distributed neural representation that is robust to the absence of the PFC contribution.

What about conjecture (2)? A strong version of it would be, to put it succinctly, that Hebb was wrong about the first of his two traces. By this account, STM would be supported by near-instantaneous modification of synaptic strength upon presentation of a to-be-remembered sample stimulus. Is this plausible? Certainly, from a computational perspective, there is no principled reason why this couldn't be so: memory can be encoded in a distributed pattern of weights in a network whether the duration of storage is 20 milliseconds, 20 seconds, or 20 years. But from a physiological perspective is there any evidence that such a "passive" (or "weight-based") mechanism may underlie some forms of STM? The answer is that, as of this writing, there is no definitive evidence of such a phenomenon *in vivo*. However, *in silico* this has been modeled as transiently elevated levels of Ca^{2+} in presynaptic terminals, which would potentiate them by increasing the amount of neurotransmitter that these terminals would release should an action potential trigger it. *In vitro*, an associative short-term potentiation that lasts on the order of 20 seconds or so has been shown to depend on a membrane-spanning subunit of the AMPA receptor, and is presumably supported by a short-lived conformational change in this protein.

A softer version of conjecture (2) is to appeal to relatively recent developments in multivariate analyses of neural signals. These call into question the validity of face-value interpretation of the information content carried in the activity of single neurons, regardless of their presumed stimulus selectivity. We will return to these the in final section of this chapter.

STM vs. working memory

Continuing with the reconciliation of the electrophysiological data summarized in *Figure 13.2* and *Figure 13.3* with the subsequent "third generation" lesion data, it is helpful to keep in mind the distinction between STM and working memory. As the excerpted quotes from Miller and colleagues make clear, their ABBA task is a test of working memory, not "just" STM, in that it requires the active comparison of each item against a stored representation of the sample, and the inhibition of the tendency to respond to just any instance of stimulus repetition. Indeed, it makes many of the same requirements as does the Self-Ordered Choosing task from Petrides (2000) that was sensitive to PFC damage. Additionally, although in *Figure 13.3* we have only highlighted the subset of PFC neurons that speak most directly to a PFC-as-working-memory-buffer idea, Miller, Erickson, and Desimone (1996) reported many other types of activity in PFC neurons. Some showed a progressive increase of activity across the trial ("climbing"), as though encoding a future event (sometimes called "prospective memory"). Some climbing neurons were also sensitive to sample identity, whereas others were not. Many more PFC than IT neurons showed match enhancement responses, and in PFC many of these did so regardless of sample identity. Finally, Miller, Erickson, and Desimone (1996) noted that "far fewer cells in PF cortex gave stimulus-selective responses than in IT cortex ... consistent with the idea that PF cortex is more involved in behavioral functions rather than visual recognition or the coding of complex objects per se" (p. 5165).

The two preceding subsections provide arguments for why the data presented in *Figure 13.2* and *Figure 13.3* may not reflect the operation of PFC-based STM buffers. *Thought Box 13.2* addresses the question, for *Figure 13.2* in particular, of *If not working memory storage, what function was this activity supporting?*

THOUGHT BOX

13.2 Complications with the interpretation of neural signals

Here, we consider three instances in which the original, seemingly straightforward, interpretation of electrophysiological findings have been called into question by subsequent studies.

First is the finding that neurons in PFC seem to be tuned for the STM of a location in space (*Figure 13.2.A*). Indeed, these neurons have been characterized as having “memory fields,” by analogy to “receptive fields” in the visual system. A study by Lebedev, Messinger, Kralik, and Wise (2004), however, has challenged this interpretation (*Figure 13.4*). Their task required the animal to keep two locations “in mind” – a remembered location and an attended one. The majority of neurons from which they recorded in the caudal half of the principle sulcus tracked the focus of attention, not the remembered location (*Figure 13.4.B*). The explanation for this finding is that many of the PFC neurons recorded by Funahashi et al. (e.g., 1989; *Figure 13.2.A*) may have been supporting attentional, rather than mnemonic, functions (*Tip 13.9*).

The second reconsidered finding is the model of a dorsal-ventral segregation of STM function in PFC (Wilson, Ó Scalaidhe, and Goldman-Rakic, 1993; *Figure 13.2.B*). This finding was challenged by Earl Miller and colleagues, in Miller’s first paper as an independent investigator, in which Rao, Rainer, and Miller (1997) trained monkeys to perform a two-step What-then-Where delayed-recognition task (*Figure 13.5.A*). They reasoned that perhaps in the earlier studies, the serial training of monkeys to perform STM for one type of information, then for another, had led their PFCs to organize themselves according to the tasks that they were trained to perform. The results indicated that over half of the PFC neurons displaying elevated delay-period activity showed tuning for both object identity and location, and that these were topographically

FIGURE 13.4 The representation of attention vs. memory in PFC. Source: Lebedev, Messinger, Kralik, and Wise, 2004. Reproduced with permission of the authors.

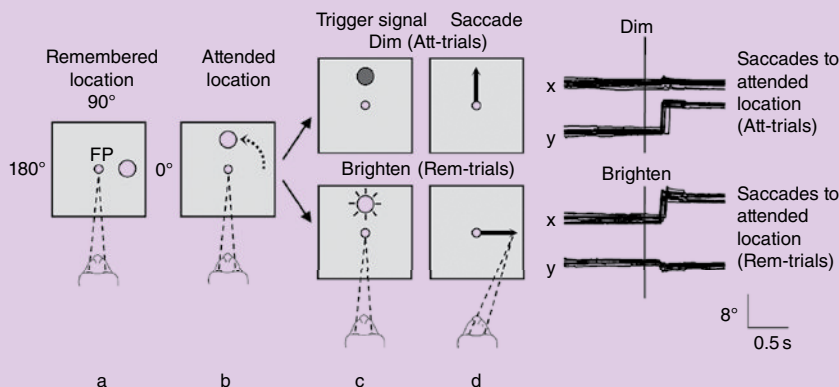


FIGURE 13.4.A Each trial began with the cuing of a “Remembered location” (at one of four possible positions). Next, the stimulus began moving along the diameter of an imaginary circle centered on the fixation point, later stopping at one of the three remaining locations (not the remembered location), thereby defining the “Attended location.” Careful attention to the stimulus was required, because the impending “Trigger signal” was very subtle: a dimming of the stimulus instructed a saccade to the stimulus itself (i.e., to the attended location), whereas a brightening instructed a saccade to the remembered location. Eye-position traces correspond to upward and rightward saccades.

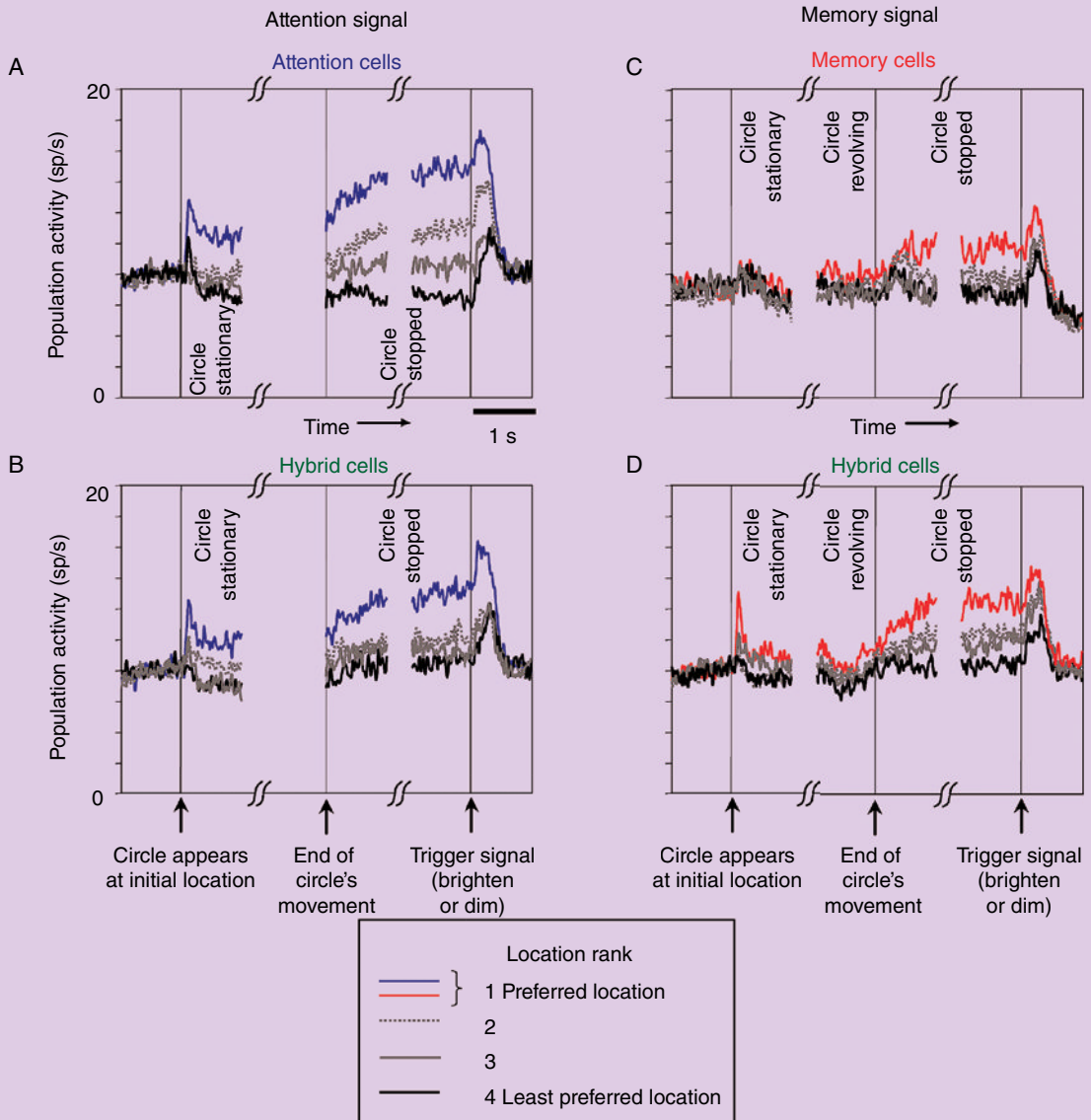
Thought Box 13.2 *Continued*

FIGURE 13.4.B Left- and right-hand columns illustrate responses on “Attention signal” and “Memory signal” trials, respectively. Across two monkeys, 61% of delay-active PFC neurons were sensitive to only the attention demands of the task, 16% to only the memory demands, and the remaining 23% to both (“Hybrid”). Plots A–D illustrate activity from these three types of neurons on the two types of trial.

Thought Box 13.2 *Continued*

FIGURE 13.5 Integration of *what* and *where* in primate prefrontal cortex. Source: Rao, Rainer, and Miller. 1997. Reproduced with permission of AAAS.

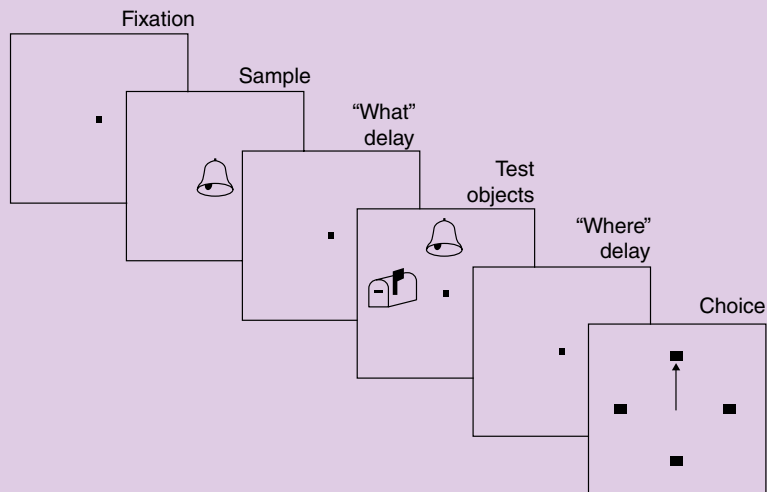


FIGURE 13.5.A An example behavioral trial. During "Test objects" epoch the animal had to implicitly select the location occupied by the sample stimulus, then, at the "Choice" epoch, make a saccade to the selected location.

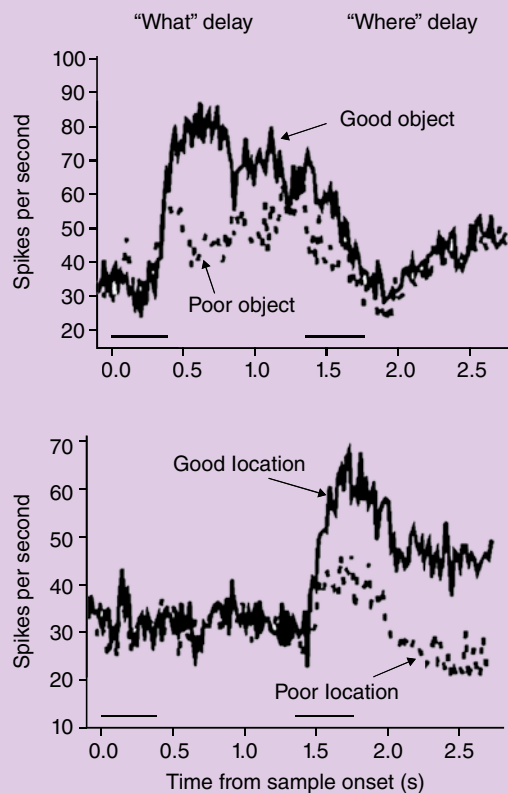


FIGURE 13.5.B Four traces from a single PFC neuron, responding to a "Good" vs. a "Poor" object (analogous to "Best" vs. "Worst" from *Figure 13.3.C*), and a "Good" vs. a "Poor" location.

Thought Box 13.2 Continued

intermingled with those that responded only to “What” or only to “Where,” with no evident dorsal-ventral segregation. This finding cast doubt on the labeled line-like organization of function in PFC suggested by *Figure 13.2.B*.

The final signal to be considered here is an ERP signal recorded during visual STM (sometimes abbreviated VSTM or VWM). This signal, most commonly referred to as the contralateral delay activity (CDA; *Tip 13.10*), has generated a great deal of excitement because it seems to track what’s happening “inside people’s heads” rather than on the surface of their retinas. Thus, if an individual’s visual STM capacity is estimated to be 3 items, the CDA will increase in magnitude, monotonically, as the memory array increases from 1 to 2 to 3 items (*Figure 13.6.A*) but here it saturates, and will remain at that magnitude for arrays of 4, 8, or 12 items, presumably because the subject can only select and remember 3 items out of an array, whether it contains 3 or 12 items. The alternative explanation, still to be proven definitively, but nonetheless provocative, derives from the fact that power in the alpha band of the EEG also increases monotonically with STM load (*Figure 13.7*). It holds that the CDA may be an artifact of oscillatory dynamics in the alpha band of the EEG (*Figure 13.8*). Were this true, fluctuations in the CDA might reflect adjustments in neural state (e.g., inhibitory tone) rather than stimulus representation, per se, in visual STM.

FIGURE 13.6 “Contralateral delay activity” (CDA), an ERP component sensitive to STM load.

Source: Vogel and Machizawa, 2004. Reproduced with permission of Nature Publishing Group.

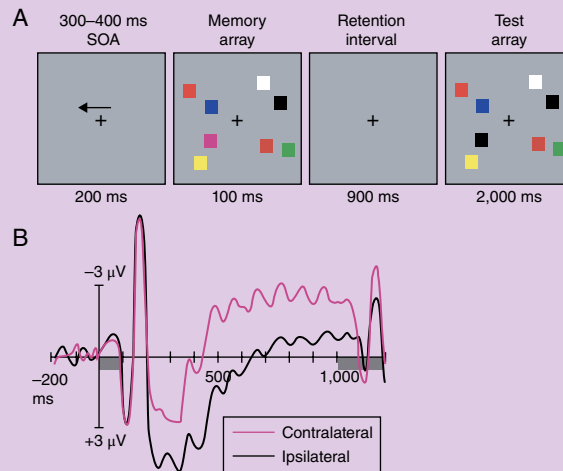


FIGURE 13.6.A Panel **A** illustrates the task, a variant of delayed recognition called a “change detection” task because an element in the cued array changes on 50% of trials. Critically, at the beginning of each trial a cue indicates whether the array in the left or right hemisphere is to be remembered. Panel **B** shows the grand average (12 subjects) ERP recorded at parietal and occipital electrodes, first averaged within each subject across electrodes from just the right and just the left hemisphere, then averaged across trials as a function of whether the attended array was in the visual field ipsilateral or contralateral to the hemisphere. The rationale is that, because subjects fixated centrally, visual stimulation for both hemispheres was comparable, and the difference in the two waveforms reflected the effect of attention and/or STM. Note that, for both waveforms, one can count ~5–6 cycles of an oscillation between time 500 and 1,000 ms, which would correspond to 10–12 Hz (i.e., the alpha band of the EEG), an observation that will be relevant when considering *Figure 13.8*.

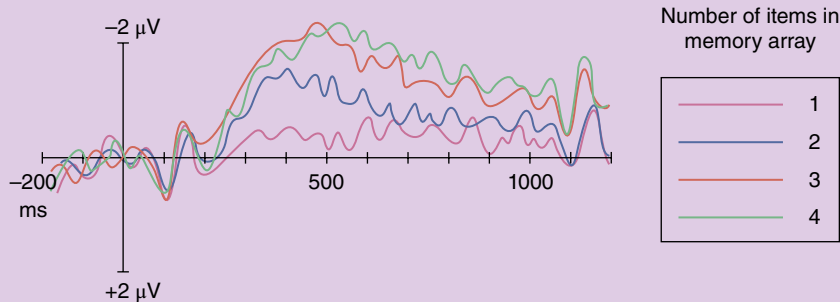
Thought Box 13.2 *Continued*

FIGURE 13.6.B The CDA at four memory loads. Beginning with load 4, the green waveform is the subtraction of the two illustrated in *Figure 13.6.A* (contralateral – ipsilateral). Note that this removes the early stimulus-evoked deflections, because these are (evidently) comparable for the two hemispheres. The remaining CDAs illustrated here are for arrays of load 1, 2, and 3. Note that the magnitude of the CDA saturates at 3, corresponding to the estimated mean visual STM capacity of this group at 2.8 items.

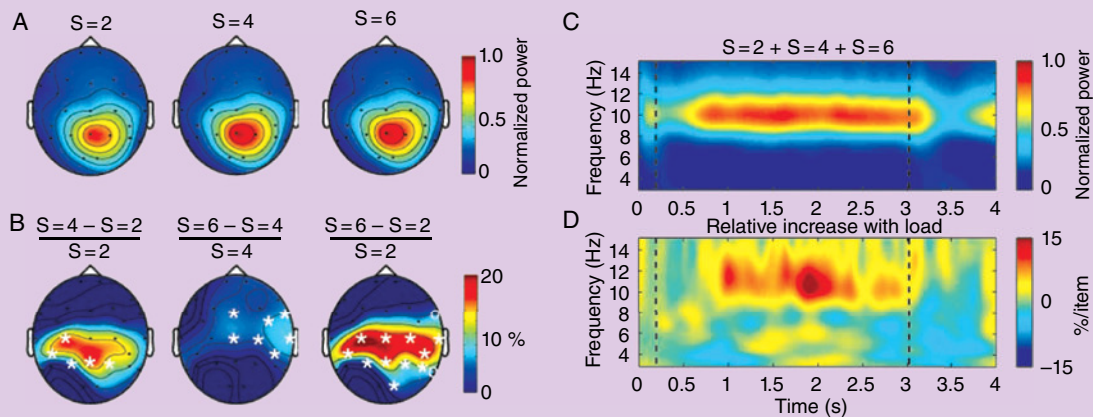


FIGURE 13.7 Alpha-band dynamics during a verbal STM task. The three topographic plots in **A** show scalp distribution of elevated delay-period alpha-band power (9–12 Hz) at memory set sizes of 2 ($S=2$), 4, and 6 consonant letters. **B** shows the changes in alpha-band power as a function of set size, white asterisks indicating electrodes where change was significant. **C** is a time-frequency spectrogram averaged across all loads and from the electrodes shown in white in panel **B**. Dashed vertical lines indicate the beginning and end of delay period. **D** shows changes in oscillatory power as a function of load. Source: Jensen, Gelfand, Kounios, and Lisman, 2002. Oxford University Press.

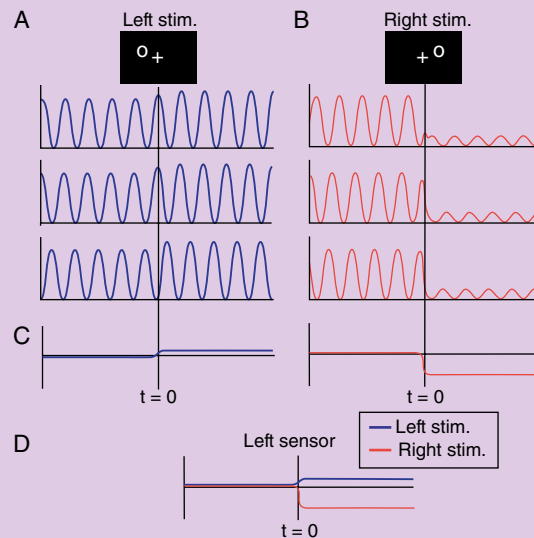
Thought Box 13.2 *Continued*

FIGURE 13.8 Hypothetical, diagrammatic explanation of how “asymmetric amplitude modulation” (AAM) could produce a CDA. **A** and **B** represent stimulus presentation to the left and right hemisphere, and the effects of each on a left hemisphere sensor (the empirical data that tested this hypothesis were acquired with MEG; the logic is equally applicable to EEG). Ipsilateral stimulus presentation produces a slight increase in the peak of the alpha-band oscillation (blue), whereas contralateral stimulus presentation produces a large decrease in the peaks (no influence on the troughs in either case). **C** shows the consequence of this AAM. In both situations, because pre-stimulus alpha-band oscillations are random in phase from trial to trial, averaging across many trials yields a flat line (i.e., they “cancel each other out”). The AAM that begins with stimulus onset ($t=0$) doesn’t change the “cancelling out” phenomenon of averaging, but has the effect of shifting the voltage of the flat line in the positive direction or the negative direction (*Tip 13.11*). If the magnitude of AAM scaled with memory load (e.g., due to increased demands on inhibition, or for more precise gating of signals), this phenomenon, when trial-averaged to produce an ERP, would result in a load-sensitive CDA. Source: van Dijk, van der Werf, Mazaheri, Medendorp, and Jensen, 2010. Reproduced with permission of PNAS Group.

Tip 13.9

Recall from *Figure 8.12* that Jerde, Merriam, Riggall, Hedges, and Curtis (2012) have made the case that, neurally, there may be no difference between remembering a location, attending to a location, or planning a saccade to a location.

Tip 13.10

Another group had previously discovered this component, and called it the *sustained posterior contralateral negativity* (SPCN), which is actually a more descriptive title than *contralateral delay activity*. They were first, and their name is better; but popularity and influence in science is capricious, and often not fairly apportioned.

Tip 13.11

Such an overall shift in the voltage of a signal is referred to as a “DC shift,” by analogy to direct current, a constant (i.e., not alternating) current source. When the magnitude of current flow changes to a new, constant value, this is a DC shift. It may be that the ECoG-measured DC shift referred to in the extended quote from Fuster and Alexander (1971; earlier in this chapter) may relate to the CDA.

A BRAVE NEW WORLD OF MULTIVARIATE DATA ANALYSIS

Your humble and devoted author has been an active researcher in the cognitive neuroscience of short-term and working memory for roughly 20 years (yikes!), and so one might think that this would be the easiest chapter for him to write. I certainly did. However, I’m now realizing that my insider status makes what I’m about to do all the more difficult, because it means leaving out so much important and influential work, by so many scientists who I really admire. (To say nothing of older work from my own career.) But here goes ...

The tradition of univariate analyses

There was a period of time, during the 1990s and 2000s, when one could be forgiven for characterizing the functional-neuroimaging-of-human-working-memory literature as a reactive one that parroted the monkey electrophysiology literature with a lag of 3–5 years. Thus, for example, a few years after the studies from Funahashi, Bruce, and Goldman-Rakic (1989) and Wilson, Ó Scalaidhe, and Goldman-Rakic (1993), neuroimaging studies appeared that seemed to confirm the gist of these findings in the human; then a few years after the Rao, Rainer, and Miller (1997) study that challenged the conclusion of Wilson, Ó Scalaidhe, and Goldman-Rakic (1993), neuroimaging studies came out challenging the conclusions of the first set of neuroimaging studies; and so on. Now, of course, this would be a cynical, and overly facile summary. Some examples of neuroimaging studies of STM making important, original contributions have

already been highlighted in, for example, *Figure 8.4* and *Figure 11.2*. Nonetheless, recent developments in the field are starting to suggest that much of the neuroimaging literature based on univariate analyses may need to be reinterpreted, for two principal reasons, both deriving from advances in multivariate pattern analysis (MVPA) of fMRI data sets. The first relates to sensitivity: MVPA from several groups has now established that regions that do not show elevated bulk signal intensity during the delay period of an STM task can nonetheless be shown to be actively representing the to-be-remembered information. The second relates to specificity: MVPA has also shown that elevated delay-period signal, even when it shows properties, like load sensitivity, that have been considered hallmarks of STM storage, doesn’t always represent the information that researchers have assumed that it does.

MVPA of fMRI

Increased sensitivity reveals delay-period storage in “silent” areas

In 2009, two groups published striking demonstrations that the active retention of information in STM can occur in regions that do not show elevated activity. One of them, from John Serences at the University of California, San Diego and his long-time collaborators from the University of Oregon (Serences, Ester, Vogel, and Awh, 2009) was designed to test the “sensory recruitment hypothesis,” the idea that the short-term retention of information may be supported, at least in part, by the very same neural systems that were responsible for their perceptual processing. (Note how this differs from the scheme, illustrated in *Figure 13.2.B*, whereby the “back of the brain” does vision and the

Tip 13.12

Note that the sensory-recruitment hypothesis would be consistent with the first of Hebb’s two hypothesized memory traces – the “reverberatory” mechanism for “a transient ‘memory’ of [a] stimulus” was proposed to occur in the very networks in which synaptic modification would encode the LTM, not in an offline STM buffer, which is the function that some had ascribed to the PFC.

“front of the brain” does memory; *Tip 13.12.*) To test it, they scanned subjects performing 10 second delayed recognition in which each sample stimulus comprised a set of black oriented bars (referred to as a grating) superimposed on a colored square, and the memory probe matched the sample along both dimensions on half the trials, differing subtly by color or by orientation in the other half. (Trials were blocked by condition.)

Figure 13.9.A shows the trial-averaged fMRI signal from the 62 most responsive voxels in V1, as defined with a (univariate) GLM analysis. (An analogy would be to Figure 8.11, which shows the spatially averaged response collapsed across the two voxels in that simulated data set.) Two things are noteworthy. First, the signal intensity drops to baseline during the delay period of both conditions. Second, it's clear by

FIGURE 13.9 MVPA reveals delay-period stimulus representation in V1. Source: Serences, Ester, Vogel, and Awh. 2009. Reproduced with permission of SAGE Publications.

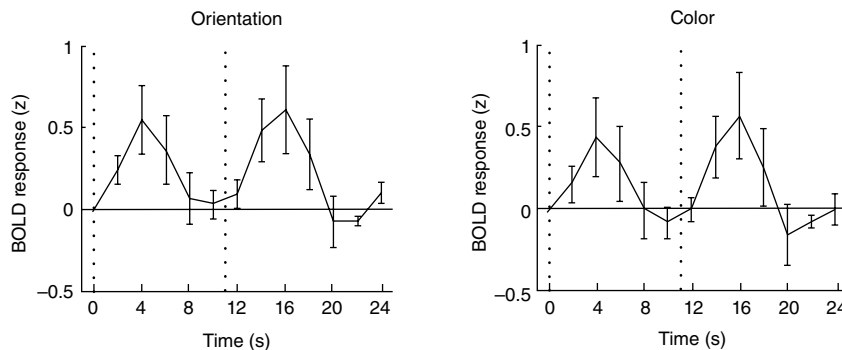


FIGURE 13.9.A Trial averaged fMRI activity from V1 during orientation and color blocks; vertical lines indicating the onset of sample and probe stimuli, respectively. Note robust stimulus-evoked responses, but that activity drops back to baseline during the delay period.

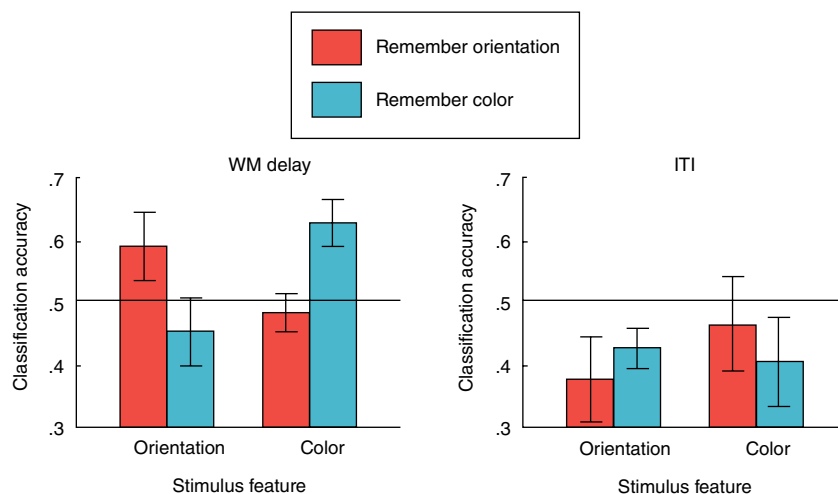


FIGURE 13.9.B Classifier performance on (or, for ITI, after) “Remember orientation” trials (red bars) and “Remember color” trials, as a function of stimulus feature being decoded. Note that the WM delay and ITI epochs were identical in length and with respect to the lag after the most recently presented stimulus (delay decoding performed on data from 4–10 seconds after presentation of the sample; ITI decoding from 4–10 seconds after presentation of the probe.)

mentally superimposing the two that one wouldn't be able to discriminate orientation from color trials with these data. Panel B of this figure, however, shows that these 62 voxels evidently carried information about the sample, because Serences et al. (2009) were able to discriminate 45°- from 135°-oriented gratings on Remember Orientation trials, and red from green on Remember Color trials.

Demonstrating stimulus-specific patterns of delay-period activity in a putatively “uninvolved” part of the brain is an impressive technical feat; a neat trick. The real conceptual advance from this study, however, was the insightful way in which the authors tested a specific prediction of the sensory recruitment hypothesis: they trained a classifier to discriminate 45°- from 135°-oriented gratings (and red from green displays) from “perception” trials – signal corresponding to the runs used to derive the ROIs, during which subjects just viewed stimuli – then used these classifiers to decode data from STM trials. Successful decoding would necessarily mean that patterns of activity during the STM trials were highly similar to patterns of activity measured during the perception trials. Thus, STM for oriented gratings (or color) could be said to engage the same neural representation as the visual perception of oriented gratings (or color). Serences et al. (2009) summarized it as so: “the sustained activation patterns observed during the delay period were similar to patterns evoked by the continuous presentation of identical sensory stimuli, which suggests that early feature-selective visual areas are recruited to maintain a ‘copy’ of remembered stimulus attributes, as opposed to a more abstract or categorical representation” (p. 213; this, and related work from this group, is featured in *Web Clip 13.1*).

It's not all peaches and cream, however. This study (and a similar one from Harrison and Tong, 2009) provides a sobering cautionary tale with regard to the hundreds (thousands?) of previously published neuroimaging studies of working memory featuring univariate analysis methods, in that many of them (including many from this author; yikes again!) may have failed to identify stimulus-related delay-period activity due to the relatively poor sensitivity of the univariate methods that we all employed up until a few years ago. This can be thought of as Type II error, which, as a general rule,

is to be minimized as much as possible. Perhaps more pernicious, though, is Type I error resulting from misinterpreting stimulus specificity.

Increased specificity reveals absence of stimulus information in elevated delay-period activity

In this section, the author will immodestly summarize research from his own group, specifically from former postdoc Steve Emrich, and graduate student Adam Riggall (degreed and gone by the time this is published?). Emrich, Riggall, LaRocque, and Postle (2013) tested STM for the direction of motion with three serially presented apertures, each containing a field of dots colored red, green, or blue. On some trials only one field of dots was moving, on some two, and on some all three. Thus, memory load was manipulated and, indeed, results of the GLM-based univariate analysis indicated that regions of rostral intraparietal sulcus (IPS) and several regions of PFC – caudal superior frontal gyrus and middle frontal gyrus, and more rostral regions of middle and inferior frontal gyri – demonstrated load sensitivity. I, myself, just a few years earlier, had written that load sensitivity is accepted as the “gold-standard” hallmark of a region's involvement in STM storage of stimulus information. MVPA, however, showed this assumption to have been flawed. As illustrated in the MVPA plots of Figure 13.10.B, regions that showed robust, elevated delay-period activity that was sensitive to load were not found to carry stimulus-related information. Although this is a null result, and we all know that the null hypothesis cannot be proven, the reason to take this null result seriously comes from the manifestly impressive sensitivity of the technique, in that it did (as have others previously) find evidence for active delay-period stimulus representation in regions whose aggregate delay-period signal intensity did not change from baseline.

And so here the cautionary tale is perhaps even more sobering than that implied in the Serences et al. (2009) and Harrison and Tong (2009) studies, in that it suggests the possibility that past interpretations of functions supported by elevated delay-period activity may need to be reconsidered; and some of these past interpretations almost assuredly will be shown to have been incorrect (*Tip 13.13*).

FIGURE 13.10 MVPA of delayed-recognition of directions of movement. Source: Emrich, Riggall, LaRocque, and Postle, 2013. Reproduced with permission of the Society of Neuroscience.

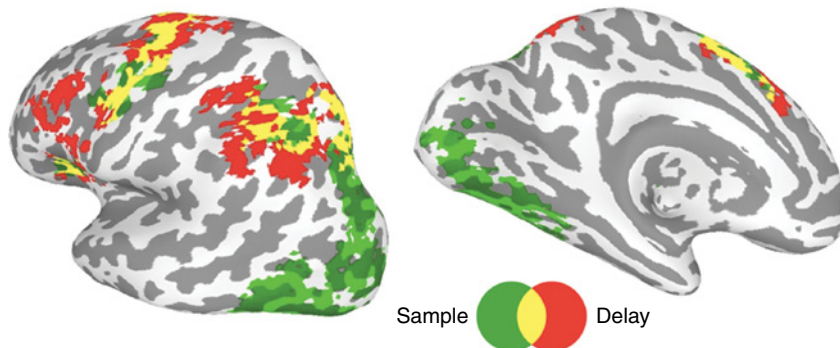


FIGURE 13.10.A Group average statistical maps, showing regions whose BOLD signal intensity level was sensitive to load during sample presentation (green), delay period (red) or both (yellow). Regions depicted in red and yellow have traditionally been considered regions supporting stimulus representation in STM.

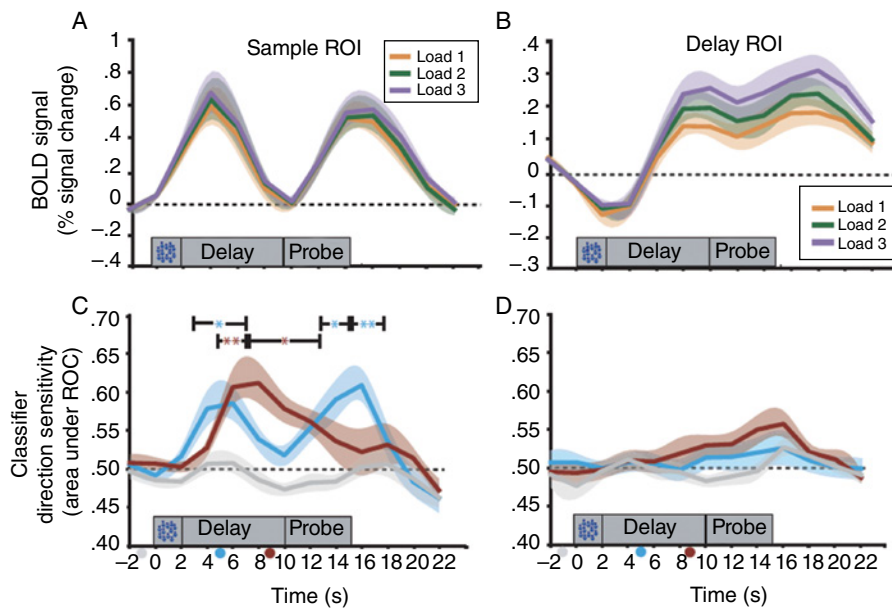


FIGURE 13.10.B Plots **A** and **B** show trial-averaged time series data, as a function of load, from regions appearing green (Sample) or red (Delay) in panel **A** (i.e., no voxels that showed both.). Note how Sample ROI plot resembles that from Serences et al. (2009; *Figure 13.9.A*). Plots **C** and **D** illustrate the decoding time course for classifiers trained at time -2 (i.e., before trial began; gray), at time 4 (i.e., maximum stimulus-evoked; teal); or at time 8 (late delay; maroon). Lack of significant decoding from delay-active voxels indicates a failure to find evidence for stimulus representation in these regions.

Tip 13.13

ERP measures such as the CDA/SPCN (*Thought Box 13.2*) are also univariate measures. What the implications of MVPA findings with fMRI may be for EEG/ERP are only now, at the time of this writing, beginning to be considered.

The neural bases of STM

The MVPA analyses from studies of fMRI data from Serences et al. (2009; *Figure 13.9*), Harrison and Tong (2009), and Emrich, Riggall, et al. (2013; *Figure 13.10*) are all consistent with the sensory-recruitment hypothesis, whereby the active retention, per se, of information in visual working memory occurs in the networks that are responsible for the perception of these stimuli, and the features that make them up. And so this is consistent with the first of Hebb's dual traces, with the caveat that this process occurs at a grain of anatomical detail that is not detectable in bulk changes of signal intensity. There are data from lesion studies and from extracellular neurophysiology that are consistent with this account, and are referenced as *Further Reading*.

And so what about the sustained, elevated activity that many studies, in monkeys and in humans, have documented in PPC and PFC? For spatial STM tasks, much of this likely reflects engagement of the dorsal frontoparietal network implicated in the endogenous control of attention (*Chapter 8*; *Figure 13.4*). For feature- and object-based STM and working memory, some insight comes from the MVPA of extracellular activity; read on!

Retrospective MVPA of single-unit extracellular recordings

In principle, the same caveats that we have just raised about to the interpretation of univariate analyses of neuroimaging data also apply to single-unit electrophysiological data. That is, face-value interpretation of elevated activity in a neuron that has been identified as "coding for" a specific kind of information can also lead to flawed inference. *Methodology Box 13.1* illustrates why, and also introduces a solution.

A population-level analysis of "Dynamic Coding for Cognitive Control in Prefrontal Cortex"

The study with which we conclude this chapter is a retrospective MVPA of a study of PFC neuronal activity in monkeys performing a delayed paired-associate recognition task (*Tip 13.14*). The first thing that one notices in *Figure 13.11.A* is the absence of elevated activity at the population level during the post-cue delay. (Also evident is a match enhancement-like response to the Target relative to Neutral and Distractor stimuli.) *Figure 13.11.B* illustrates that although population level of activity is not elevated during the delay period, this population of PFC neurons is nonetheless representing the trial context. Are these neurons holding a retrospective representation of the trial-initiating cue? Or a prospective representation of that cue's associate (i.e., the Target for that trial)? To explore this, Stokes et al. (2013) applied the following reasoning, and analytic procedures:

The activation state of the full population was represented as a 627-dimensional coordinate in Euclidean space, where each dimension represents the instantaneous firing rate of a single neuron estimated within 50 ms sliding windows. The dynamic trajectory through this state space is the path that passes through the multidimensional coordinate of each time point. To explore the dynamic behavior of this network, we ... calculated the Euclidean distance between trajectory pairs ... as a function of time. Distances between all contributing pairs of conditions (e.g., all pairs of cues) were then averaged to yield a single

Tip 13.14

The senior author on this study, John Duncan, holds a place in our "Pantheon of Professional Courage" (joining, so far, Nancy Kanwisher (*Chapter 9*) and Martha Farah (*Chapter 9*)). Although there are many researchers who started their careers as cognitive psychologists and gradually added neuroimaging to their research program, Duncan is the only one with whom I am familiar who, mid-career, launched an awake, behaving monkey laboratory to complement his already-groundbreaking research in cognitive psychology and cognitive neuroimaging.

METHODOLOGY BOX

13.1 Retrospective MVPA of single-unit extracellular recordings

The traditional approach to in vivo extracellular electrophysiology is to begin each day's recording session by inserting an electrode into the microdrive, positioning it over the region of cortex that is of interest, and lowering it into the brain until a neuron is isolated. It's painstaking work, and, when training the animal(s) is taken into account, it can take considerably longer than a year (sometimes many years) to acquire a complete set of data. In terms of the investment of blood, sweat, and tears – not to mention the PI's research funding – these data sets are priceless.

The rationale behind multivariate analyses, however, is that the system (i.e., the brain) experiences the simultaneous activity of many, many processing elements all at the same time (e.g., the 62 voxels in the Serences et al. (2009) V1 ROI, or the hundreds of “sensory” vs. “delay” voxels in the Emrich, Riggall, et al. (2013) study), and that the overall influence of these patterns of activity is very likely to be different from what one could deduce by studying any single processing element in isolation. What is a neurophysiologist to do?

One approach that has yielded important insights is to retrospectively assemble all of the records of single neurons recorded while the animal performed identical trials, and analyze them all together as though all had been recorded simultaneously during a single “virtual trial.” Pioneers in this approach have been Ethan Meyers and his mentor Tomaso Poggio at MIT. In one study, they reanalyzed a data set from Earl Miller's laboratory in which monkeys had viewed images of 3-D computer-generated animals and judged whether each was more cat-like or dog-like. The original study had compared neuronal activity in PFC vs. IT, finding that IT seemed “more involved in the analysis of currently viewed shapes, whereas the PFC showed stronger category signals, memory effects, and a greater tendency to encode information in terms of its behavioral meaning” (Freedman, Riesenhuber, Poggio, and Miller, 2003, p. 5235). The multivariate reanalysis, although generally confirming these broader patterns, uncovered novel, surprising information about the interpretation of sustained, elevated firing rates: the representation of stimulus category information in a delayed-match-to-category task “is coded by a nonstationary pattern of activity that changes over the course of a trial with individual neurons ... containing information on much shorter time scales than the population as a whole” (Meyers, Freedman, Kreiman, Miller, and Poggio, 2008, p. 1407). That is, this information was not carried for extended periods in time by any individual neuron, regardless of whether or not its activity level was elevated at a sustained level across the delay period.

Whoa! This observation, if true, would have major implications for how one interprets neuronal recordings. And lest one think that this may have been a spurious finding attributable to the retrospective nature of the analyses, a group led by (former Goldman-Rakic trainee) Matthew Chafee at the University of Minnesota has reported an analogous phenomenon from simultaneous (i.e., “prospective”) recordings acquired with multiunit arrays recording from PPC, illustrating that a *stable* mental representation of a part of an object is sustained by a succession of activity patterns that change on a 100 ms by 100 ms time scale.

summary statistic for each time point ... This distance measure is closely related to the population coding metric [because] accurate pattern decoding critically depends on a reliable multidimensional distance between conditions. Position in state space was visualized using

classical MDS [multidimensional scaling; a technique for displaying high-dimensional data in 2-D]. For each analysis, we made four independent estimates of the population activity for each condition of interest by averaging every fourth trial within that condition. MDS was then

FIGURE 13.11 MVPA of PFC activity during a working memory task. Source: Stokes, Kusunoki, Sigala, Nili, Gaffan, and Duncan, 2013. Reproduced with permission of Elsevier.

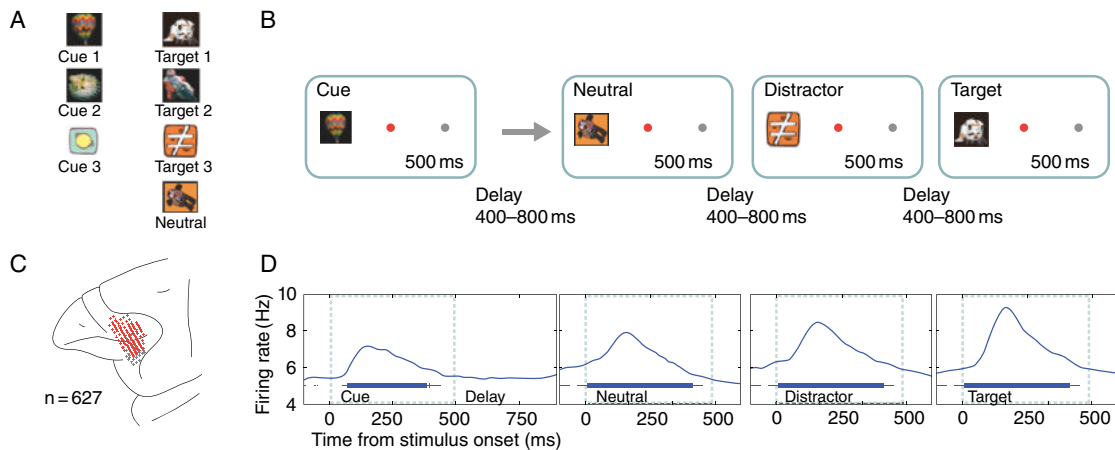


FIGURE 13.11.A Panel **A** illustrates the arbitrarily determined associations that monkeys learned prior to the working memory task. **B** illustrates an example trial: not unlike ABBA, the trial-initiating cue informs the animal which stimulus it should respond to. Note that, on other trials, the stimulus labeled “Target 1” could serve as a Distractor and the stimulus labeled “Target 3” could serve as the Target. Thus, on each trial, a target’s behavioral relevance is conditioned on the Cue. **C** illustrates, with color coding, recording sites for the two monkeys from which 627 neurons were recorded. **D** illustrates the mean firing rate of the population, with the dim dashed box indicating the period of stimulus presentation, and the thin and thick blue horizontal lines indicating periods during which the population activity was elevated relative to baseline (at $p < 0.05$ and < 0.0001 , respectively).

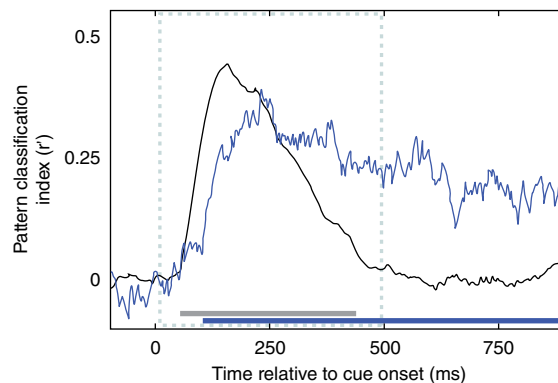


FIGURE 13.11.B Superposition of cue-related mean firing rate (gray, represented from Figure 13.11.A) and performance of classifier at discriminating between the three trial types. Gray bar indicates significantly elevated activity vs. baseline, blue bar indicates significant classifier performance. (Classifier performance is expressed as a transformed correlation coefficient indexing how closely trained patterns match decoded patterns.)

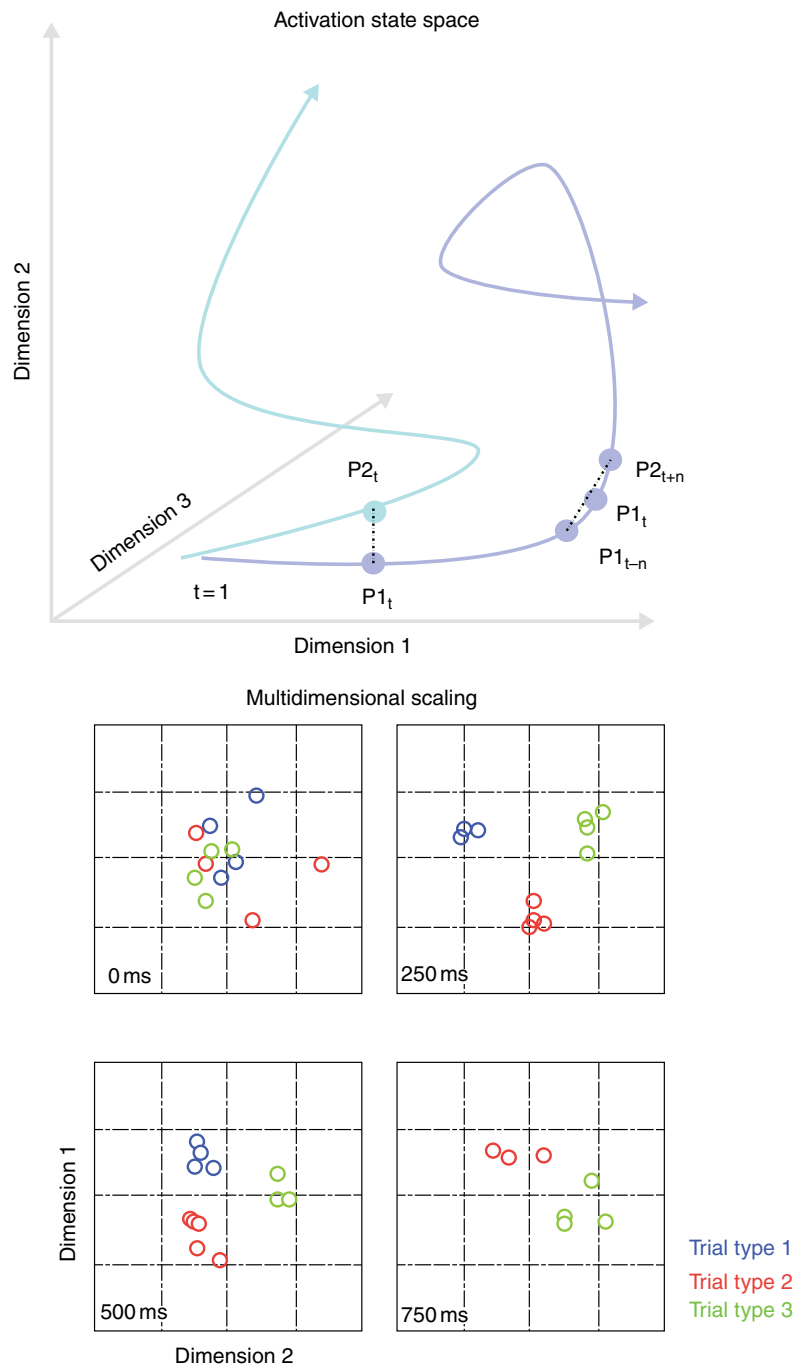


FIGURE 13.11.C The top plot is a cartoon of two hypothetical trajectories of PFC activity during two different behavioral conditions (here labeled P1 and P2; we can think of them as Cue 1 and Cue 2). The dotted line between $P1_t$ and $P2_t$ illustrates the “Euclidean distance between trajectory pairs” referred to in the quoted text from Stokes et al. (2013, p. 373). The dotted line along $P1_{t-n} - P1_t - P1_{t+n}$ corresponds to the position of PFC activity in this high-dimensional state space at three points in time. The four plots labeled “Multidimensional scaling” illustrate the calculated Euclidean distance between trial types at four points in time relative to cue onset. Within each plot, each of the four identically colored circles corresponds to one of four independent estimates calculated for that trial type.

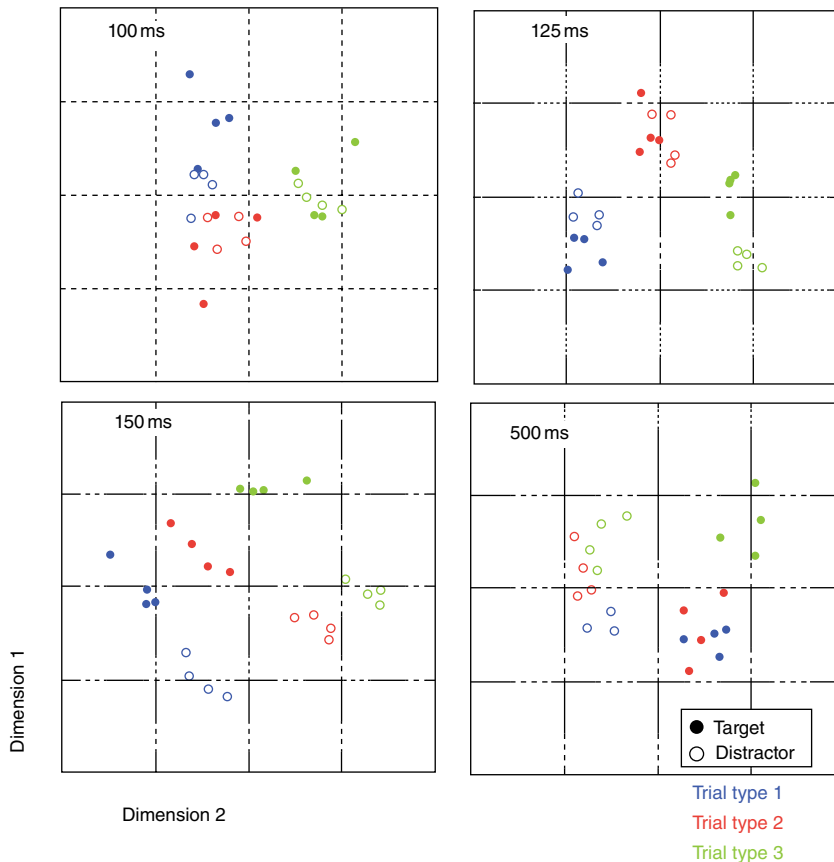


FIGURE 13.11.D In contrast to *Figure 13.11.C*, which features the trial-initiating cue, this figure presents an MDS depiction of the evolving PFC representation of each associate (differentiated by color) as a function of behavioral context (*On this trial, does the stimulus serve as a target or as a distractor?*; differentiated by fill). Over the course of the 500 ms that the associate is on the screen, the PFC transitions from representing its identity (trial types clustered by color) to representing response context (trial types clustered by whether or not stimulus requires a response).

performed on the Euclidean distance matrix. Data were plotted against the first two dimensions. The distribution of the four within condition data points provides a sense of the tightness of condition-dependent clustering. (p. 373)

The top plot in *Figure 13.11.C* is a simplified cartoon of the phenomenon measured by Stokes et al. (2013) in that they tracked the trajectory of PFC through a 627-D space (here cartooned as 3-D). The actual trajectories that they mapped are illustrated in the bottom plots, in which, through the magic of MDS, locations in the 627-D are collapsed down to 2-D plots that retain a representation of Euclidean

distance. Up to this point, what this has shown is a fancy, “under-the-hood” view of the high-dimensional structure that MVPA uses to classify delay-period activity after the presentation of a sample or cue stimulus. Even more tantalizing is the analysis, illustrated in *Figure 13.11.D*, showing how PFC processes each of the three associates, both on trials when they serve as a Target and on trials when they serve as a Distractor. (The three stimuli associated with Cues 1, 2, and 3 are labeled “Target 1, Target 2, and Target 3” in *Figure 13.11.A*, but this is confusing because each can either serve as a Target or as a Distractor; so here we will refer to them as “associates.”) What *Figure*

13.11.D illustrates a transition in PFC from stimulus-specific coding (i.e., representing Associate 1, 2, or 3) to context-dependent coding for the behavioral decision (i.e., *Is this a Target, which requires a “go,” or a Distractor, which requires a “no-go”?*). Between time 100–125 ms we see that the PFC first configures itself to represent stimulus identity (i.e., all the reds, blues, and greens cluster, regardless of their status as Target or Distractor). Twenty-five ms later, we see that PFC is representing both stimulus identity (clustering by colors) and behavioral decision (separation of Targets from Distractors). By time 500 ms (the deadline by which the animal has to respond to Targets), PFC is representing behavioral decision, but no longer stimulus identity. (Detailed consideration of this study is also presented in *Web Clip 13.2*.)

THREE-QUARTERS OF A CENTURY

We’ve covered research spanning from 1936 to 2013, from massive bilateral PFC lesions to exquisitely sensitive analysis of PFC neurophysiology carried out in 627-D space. Over the course of those roughly 75 years, we have learned a great deal about how people hold information in mind for short periods of time in order to guide their behavior. One of the themes that we’ve emphasized – that there’s an important difference between “just” holding information in mind (STM) and performing various mental operations with, or on, that information (working memory) – offers a nice segue into our next chapter, because working memory is one example of cognitive control.

END-OF-CHAPTER QUESTIONS

1. What is (are) the difference(s) between STM and working memory? Why is it important for one to be mindful of these differences when designing or interpreting the results from a study?
2. Were the interpretations that Fuster and Alexander (1971) and that Kubota and Niki (1971) gave to their electrophysiological findings consistent with the lesion data that had come before? How, or how not?
3. How did the development of the multicomponent model of working memory, from cognitive psychology, influence some people’s interpretation of the findings from question 2?
4. Summarize the findings from Funahashi, Bruce, and Goldman-Rakic (1989) and Wilson, Ó Scalaidhe, and Goldman-Rakic (1993), and how subsequent studies have been designed to evaluate and, in some cases, reinterpret them.
5. What is the CDA (or SPCN)? What distinctive property makes it particularly interesting to students of visual STM?
6. In what ways have recent MVPA analyses of fMRI studies of STM raised questions about the interpretation of earlier studies analyzed with univariate methods? (Hint: sensitivity; specificity.)
7. What new insights have come from MVPA of extracellular neuronal recordings?
8. What is our best, current understanding of the neural bases of STM? Of working memory?

REFERENCES

- Baddeley, Alan D., and Graham Hitch. 1974. “Working Memory.” In *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 8, edited by Gordon H. Bower, 47–89. New York: Academic Press.
- Emrich, Stephen M., Adam C. Riggall, Joshua J. LaRocque, and Bradley R. Postle. 2013. “Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory.” *Journal of Neuroscience* 33 (15): 6516–6523. doi: 10.1523/JNEUROSCI.5732-12.2013.
- Freedman, David J., Maximilian Riesenhuber, Tomaso Poggio, and Earl K. Miller. 2003. “A Comparison of Primate

- Prefrontal and Inferior Temporal Cortices during Visual Categorization." *Journal of Neuroscience* 23 (12): 5235–5246.
- Funahashi, Shintaro, Charles J. Bruce, and Patricia S. Goldman-Rakic. 1989. "Mnemonic Coding of Visual Space in the Monkey's Dorsolateral Prefrontal Cortex." *Journal of Neurophysiology* 61 (2): 331–349.
- Fuster, Joaquin M., and Garrett E. Alexander. 1971. "Neuron Activity Related to Short-Term Memory." *Science* 173 (3997): 652–654. doi: 10.1126/science.173.3997.652.
- Ghent, Lila, Mortimer Mishkin, and Hans-Lukas Teuber. 1962. "Short-term Memory after Frontal-Lobe Injury in Man." *Journal of Comparative and Physiological Psychology* 55 (5): 705–709. doi: 10.1037/h0047520.
- Harrison, Stephenie A., and Frank Tong. 2009. "Decoding Reveals the Contents of Visual Working Memory in Early Visual Areas." *Nature* 458 (7238): 632–635. doi: 10.1038/nature07832.
- Hebb, Donald Olding. 1949. *The Organization of Behavior; A Neuropsychological Theory*. New York: Wiley.
- Jacobsen, C. F. 1936. "The Functions of the Frontal Association Areas in Monkeys." *Comparative Psychology Monographs* 13: 1–60.
- Jensen, Ole, Jack Gelfand, John Kounios, and John E. Lisman. 2002. "Oscillations in the Alpha Band (9–12 Hz) Increase with Memory Load During Retention in a Short-Term Memory Task." *Cerebral Cortex* 12 (8): 877–882. doi: 10.1093/cercor/12.8.877.
- Jerde, Trenton A., Elisha P. Merriam, Adam C. Riggall, James H. Hedges, and Clayton E. Curtis. 2012. "Prioritized Maps of Space in Human Frontoparietal Cortex." *Journal of Neuroscience* 32 (48): 17382–17390. doi: 10.1523/JNEUROSCI.3810-12.2012.
- Kubota, Kisou, and Hiroaki Niki. 1971. "Prefrontal Cortical Unit Activity and Delayed Alternation Performance in Monkeys." *Journal of Neurophysiology* 34 (3): 337–347.
- Lebedev, Mikhail A., Adam Messinger, Jerald D. Kralik, and Steven P. Wise. 2004. "Representation of Attended Versus Remembered Locations in Prefrontal Cortex." *PLoS Biology* 2 (11): e365. doi: 10.1371/journal.pbio.0020365.
- Malmo, Robert B. 1942. "Interference Factors in Delayed Response in Monkeys after Removal of Frontal Lobes." *Journal of Neurophysiology* 5 (4): 295–308.
- Meyers, Ethan M., David J. Freedman, Gabriel Kreiman, Earl K. Miller, and Tomaso Poggio. 2008. "Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex." *Journal of Neurophysiology* 100 (3): 1407–1419. doi: 10.1152/jn.90248.2008.
- Miller, Earl K., and Robert Desimone. 1994. "Parallel Neuronal Mechanisms for Short-Term Memory." *Science* 263 (5146): 630–632. doi: 10.1126/science.8290960.
- Miller, Earl K., Cynthia A. Erickson, and Robert Desimone. 1996. "Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque." *Journal of Neuroscience* 16 (16): 5154–5167.
- Miller, Earl K., Lin Li, and Robert Desimone. 1991. "A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex." *Science* 254 (5036): 1377–1379. doi: 10.1126/science.1962197.
- Milner, Brenda. 1964. "Some Effects of Frontal Lobectomy in Man." In *The Frontal Granular Cortex and Behavior*, edited by John M. Warren and Konrad Akert, 313–334. New York: McGraw-Hill.
- Petrides, Michael. 2000. "Dissociable Roles of Mid-Dorsolateral Prefrontal and Anterior Inferotemporal Cortex in Visual Working Memory." *Journal of Neuroscience* 20 (19): 7496–7503.
- Postle, Bradley R. 2009. "The Hippocampus, Memory, and Consciousness." In *Neurology of Consciousness*, edited by Steven Laureys and Giulio Tononi, 326–338. Amsterdam: Elsevier.
- Pribram, Karl H., Albert J. Ahumada Jr., J. Hartog, and L. Roos. 1964. "A Progress Report on the Neurological Processes Distributed by Frontal Lesions in Primates." In *The Frontal Granular Cortex and Behavior*, edited by John M. Warren and Konrad Akert, 28–55. New York: McGraw-Hill.
- Rao, S. Chenchal, Gregor Rainer, and Earl K. Miller. 1997. "Integration of What and Where in the Primate Prefrontal Cortex." *Science* 276 (5313): 821–824. doi: 10.1126/science.276.5313.821.
- Serences, John T., Edward F. Ester, Edward K. Vogel, and Edward Awh. 2009. "Stimulus-Specific Delay Activity in Human Primary Visual Cortex." *Psychological Science* 20(2):207–214. doi:10.1111/j.1467-9280.2009.02276.x.
- Stokes, Mark G., Makoto Kusunoki, Natasha Sigala, Hamed Nili, David Gaffan, and John Duncan. 2013. "Dynamic Coding for Cognitive Control in Prefrontal Cortex." *Neuron* 78 (2): 364–375. doi: 10.1016/j.neuron.2013.01.039.
- van Dijk, Hanneke, Jurrian van der Werf, Ali Mazaheri, W. Pieter Medendorp, and Ole Jensen. 2010. "Modulations in Oscillatory Activity with Amplitude Asymmetry can Produce Cognitively Relevant Event-Related Responses." *Proceedings of the National Academy of Sciences of the United States of America* 107 (2): 900–905. doi: 10.1073/pnas.0908821107.
- Vogel, Edward K., and Maro G. Machizawa. 2004. "Neural Activity Predicts Individual Differences in Visual Working Memory Capacity." *Nature* 428 (6984): 679–782. doi: 10.1038/nature02447.
- Warren, John M., and Konrad Akert, eds. 1964. *The Frontal Granular Cortex and Behavior*. New York: McGraw-Hill.
- Wilson, Fraser A. W., Séamas P. Ó Scalaidhe, and Patricia S. Goldman-Rakic. 1993. "Dissociation of Object and Spatial Processing Domains in Primate Prefrontal Cortex." *Science* 260 (5116): 1955–1958.

OTHER SOURCES USED

- Crowe, David A., Bruno B. Averbeck, and Matthew V. Chafee. 2010. "Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex." *Journal of Neuroscience* 30 (35): 11640–11653. doi: 10.1523/JNEUROSCI.0954-10.2010.
- Erickson, Martha A., Lauren A. Maramba, and John Lisman. 2010. "A Single Brief Burst induces GluR1-Dependent Associative Short-Term Potentiation: A Potential Mechanism for Short-Term Memory." *Journal of Cognitive Neuroscience* 22 (11): 2530–2540. doi: 10.1162/jocn.2009.21375.
- Funahashi, Shintaro, Charles J. Bruce, and Patricia S. Goldman-Rakic. 1990. "Visuospatial Coding in Primate Prefrontal Neurons Revealed by Oculomotor Paradigms." *Journal of Neurophysiology* 63 (4): 814–831.
- Funahashi, Shintaro, Matthew V. Chafee, and Patricia S. Goldman-Rakic. 1993. "Prefrontal Neuronal Activity in Rhesus Monkeys Performing a Delayed Anti-Saccade Task." *Nature* 365 (6448): 753–756. doi: 10.1038/365753a0.
- Miller, George Armitage. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (2): 81–97. doi: 10.1037/h0043158.
- Miller, George Armitage, Eugene Galanter, and Karl H. Pribram. 1960. *Plans and the Structure of Behavior*. New York: Holt.
- Mongillo, Gianluigi, Omri Barak, and Misha Tsodyks. 2008. "Synaptic Theory of Working Memory." *Science* 319 (5869): 1543–1546. doi: 10.1126/science.1150769.
- Nakamura, Katsuki, and Kisou Kubota. 1995. "Mnemonic Firing of Neurons in the Monkey Temporal Pole During a Visual Recognition Memory Task." *Journal of Neurophysiology* 74 (1): 162–178.
- Niki, Hiroaki. 1974. "Differential Activity of Prefrontal Units During Right and Left Delayed Response Trials." *Brain Research* 70 (2): 346–349. doi: 10.1016/0006-8993(74)90324-2.
- Niki, Hiroaki, and Masataka Watanabe. 1976. "Prefrontal Unit Activity and Delayed Response: Relation to Cue Location Versus Direction of Response." *Brain Research* 105 (1): 79–88. doi: 10.1016/0006-8993(76)90924-0.
- Riggall, Adam C., and Bradley R. Postle. 2012. "The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging." *Journal of Neuroscience* 32 (38): 12990–12998. doi: 10.1523/JNEUROSCI.1892-12.2012.
- Rushworth, Matthew F. S., Philip D. Nixon, Madeline J. Eacott, and Richard E. Passingham. 1997. "Ventral Prefrontal Cortex Is Not Essential for Working Memory." *Journal of Neuroscience* 17 (12): 4829–4838.
- Suzuki, Wendy A., Earl K. Miller, and Robert Desimone. 1997. "Object and Place Memory in the Macaque Entorhinal Cortex." *Journal of Neurophysiology* 78 (2): 1062–1081.

FURTHER READING

- Mazaheri, Ali, and Ole Jensen. 2008. "Asymmetric Amplitude Modulations of Brain Oscillations Generate Slow Evokes Responses." *Journal of Neuroscience* 28 (31): 7781–7787. doi: 10.1523/JNEUROSCI.1631-08.2008.
The mathematics, and the rationale, behind calculating the AAM that may underlie the CDA/SPCN.
- Meyers, Ethan, and Gabriel Kreiman. 2011. "Tutorial on Pattern Classification in Cell Recording." In *Visual Population Codes: Toward a Common Multivariate Framework for Cell Recording and Functional Imaging*, edited by Nikolaus Kriegeskorte and Gabriel Kreiman, 517–538. Cambridge, MA: MIT Press.
Details on Meyers' and Kreiman's method for retrospective MVPA of single-unit data sets.
- Pasternak, Tatiana, and Mark W. Greenlee. 2005. "Working Memory in Primate Sensory Systems." *Nature Reviews Neuroscience* 6 (2): 97–107. doi: 10.1038/nrn1603.
Extensive review of evidence, from monkey neurophysiology and human psychophysics, for the sensory-recruitment hypothesis of STM.
- Postle, Bradley R. 2006. "Working Memory As an Emergent Property of the Mind and Brain." *Journal of Neuroscience* 139 (1): 23–38. doi: 10.1016/j.neuroscience.2005.06.005.
Review covering much of the univariate analysis-based neuroimaging literature on human short-term and working memory that this chapter skips over.

SECTION IV

HIGH-LEVEL

COGNITION

INTRODUCTION TO SECTION IV HIGH-LEVEL COGNITION

Who's to say what's "high-level cognition" vs. "mental representation"? Isn't sorting out whether it was Brazil or Uruguay who won the first World Cup, or whether it was Graham Greene or Evelyn Waugh who wrote *Our Man in Havana*, high-level cognition? Sure. But they're also examples of semantic cognition. Isn't the endogenous control of attention a form of cognitive control? Of course it is. So is, as we have noted, working memory. The labels aren't so important. The rationale behind this book's organization is that *Section II* provided a necessary foundation for exploring different facets of mental representation (i.e., *Section III*) in a substantive way, and *Section II* and *Section III* have provided

the background and framework needed for us to do the same with the topics to be covered here. None of them won't require a good understanding of perception, of attention, of motor control, and of memory. (Nor, for that matter, of oscillatory coherence, of fMRI, of neuropsychology, of computational modeling ...) Further, the chapters in this section will build on each other: principles from cognitive control will apply to decision making, emotions, social behavior, and language; principles from decision making will apply to emotions and social behavior; and so on.

It's THE NEURAL BASES OF HIGH-LEVEL COGNITION, people! Come on!

CHAPTER 14

COGNITIVE CONTROL

KEY THEMES

- Cognitive control refers to the ability to guide behavior and prioritize information processing in accordance with internal goals, regardless of the salience of goal-irrelevant stimuli in the environment, or of goal-irrelevant behavioral choices.
- The environmental-dependency syndrome resulting from lateral PFC damage produces behavioral disinhibition that is a dramatic illustration of failures of cognitive control.
- A cardinal laboratory-based operationalization of impaired cognitive control is perseverative responding on the Wisconsin Card-Sorting Test.
- ERPs from patients with lateral PFC lesions who are performing a STM task reveal two distinctive patterns: an abnormally reduced N1 to the target stimulus, suggesting impaired attentional selection/encoding; and abnormally large medium-latency AEPs to delay-period distracting stimuli, suggesting impaired filtering of task-irrelevant information.
- A computational simulation of the A-not-B error, a perseverative response in infants attributed to an immature PFC, suggests that it may arise from ineffective top-down selection of the weaker, but behaviorally appropriate, response, rather than a failure to inhibit the stronger, but inappropriate response.
- The PFC has unique properties that may underlie its central role in cognitive control: the ability to transition between an “up” state that sustains an active pattern of activity and a “down” state during which it can update to a new pattern; heavy innervation by midbrain dopamine neurons that carry reward-related information; and extensive anatomical connectivity through which it can implement control.
- The dynamics of ascending dopamine signals, including the reward prediction error signal, are of central importance for triggering state transitions in PFC and providing performance-related information to performance-monitoring circuits in the dorsomedial frontal cortex.
- The implementation of cognitive control entails prioritizing information processing in selected channels via interareal oscillatory synchrony that is, in some cases, coordinated by the thalamus.

CONTENTS

KEY THEMES

THE LATERAL FRONTAL-LOBE SYNDROME

Environmental dependency syndrome
Perseveration
Electrophysiology of the frontal-lobe syndrome
Integration?

MODELS OF COGNITIVE CONTROL

Developmental cognitive neuroscience
The A-not-B error
Is the A-not-B error a failure of working memory?
Generalizing beyond development
What makes the PFC special?
Properties of the PFC
Influence of the DA reward signal on the functions of PFC

Combining a reward prediction error signal with a flexible PFC to overcome perseveration

NEURAL ACTIVITY RELATING TO COGNITIVE CONTROL

Error monitoring
The ERN/Ne
Performance monitoring and the dorsomedial frontal cortex
Oscillatory synchrony
Thalamic control of cortical synchrony

WHERE IS THE CONTROLLER?

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

Cognitive control governs much of our lives, much of the time, often in seemingly trivial ways that nonetheless add up to us conducting our daily lives much more efficiently and efficaciously. Often we are not aware of its operation until it fails. A classic example for this writer is stepping onto the elevator in the psychology building, lifting his eyes off the document he's reading long enough to press "5," then going back to the document. The elevator rises, stops, the doors open, and he steps out. Only to find himself on the fourth floor. Here is how two leading researchers (one of them Earl Miller, from the previous chapter) introduced a very influential paper on the topic:

One of the fundamental mysteries of neuroscience is how coordinated, purposeful behavior arises from the distributed activity of billions of neurons in the brain. Simple behaviors can rely on relatively straightforward interactions between the brain's input and output systems. Animals with fewer than a hundred thousand neurons (in the human brain there are 100 billion or more neurons) can approach food and avoid predators. For animals with larger brains, behavior is more flexible. But flexibility carries a cost: although our elaborate sensory and motor systems provide detailed information about the external world and make available a large repertoire of actions, this introduces greater potential for interference and confusion. The richer information we have about the world and the greater number of options for behavior require appropriate attentional, decision-making, and coordinative functions, lest uncertainty prevail. To deal with this multitude of possibilities and to curtail confusion, we have evolved mechanisms that coordinate lower-level sensory and motor processes along a common theme, an internal goal. (Miller and Cohen, 2001, pp. 167–168)

In the elevator anecdote, the internal goal of our protagonist was to get to his office (located on the fifth floor), the detailed sensory information that was driving his behavior was being displayed on the electronic tablet in his hand. (The ubiquity of "smart phones" and other mobile devices in current society has made the value of cognitive control in everyday life all the more apparent.) As this anecdote suggests, cognitive control often implicates the idea of inhibition – the absent-minded professor's attention focused elsewhere, he failed to inhibit the prepotent reflex of walking through the just-opened doors.

Whether or not control truly engages inhibition, at a mechanistic level, will be a theme that will recur here, and in subsequent chapters.

As with many other domains of human behavior, behavioral neurology has provided a rich set of cases, and theories, that provide a backdrop for contemporary thinking about cognitive control.

THE LATERAL FRONTAL-LOBE SYNDROME

In reproducing the quote from Miller and Cohen (2001), above, I left out the following, because I want to be careful not to advocate a "one-brain-area = one-kind-of-behavior" style of thinking. To exclude it altogether, however, would be to mischaracterize the received wisdom on this subject: "This ability for cognitive control no doubt involves neural circuitry that extends over much of the brain, but it is commonly held that the prefrontal cortex (PFC) is particularly important" (p. 168). This common holding derives, in part, from a large body of undeniably compelling evidence from behavioral neurology. The French neurologist François Lhermitte (1921–1998) captured this succinctly in companion papers entitled *Human Autonomy and the Frontal Lobes* (Tip 14.1). "The patients' behavior was striking," he writes, "as though implicit in the environment was an order to respond to the situation in which they found themselves" (Lhermitte, Pillon, and Serdaru, 1986, p. 335). "*Human autonomy*"?!? *I thought this book was about neuroscience. This isn't a trick*

Tip 14.1

Often one finds terms like "frontal-lobe syndrome," "frontal-lobe damage," and, indeed, "prefrontal cortex" used without further specification of what part of the PFC is implicated. Unless otherwise stated, the focus of this chapter will be on lateral PFC. We'll see in the subsequent chapters on decision making and emotion that orbital and medial PFC, associated with a different cortico-basal ganglia-thalamic loop (Figure 7.3), subserve very different functions than lateral PFC.

to lure us into reading a treatise on existentialism, is it? Well, no. On the other hand, from here on out there's not a topic that we'll consider without clear implications for what it is that makes us uniquely human.

Environmental dependency syndrome

Building on the well-characterized phenomenon of **utilization behavior** – an exaggerated tendency for one's behavior to be determined by the external environment – Lhermitte, Pillon, and Serdaru (1986) carried out a study in which the experimenter simply sat opposite each subject and, with a “completely neutral and indifferent” demeanor, performed a series of gestures: common body gestures (e.g., crossing legs, tapping fingers); symbolic gestures (e.g., thumbing one's nose, military salute); manipulating objects (folding a piece of paper and putting it into an envelope, combing hair); and so on. No advance explanation or instructions were provided. The description of the response of neurologically intact individuals is not only hilarious, but also provides context that helps one appreciate just how striking, and profound, are the impairments of the patients. And why Lhermitte invokes autonomy in relation to the frontal-lobe syndrome:

Normal subjects never imitated the examiner. They were unconcerned but surprised, without otherwise making the slightest remark When asked why the examiner had performed these gestures, they hesitated and replied: “To test me”; “I don't know – perhaps to see my reactions”; and so on. The answers varied little with personality or age. When the examiner asked them if it had crossed their mind to imitate him, their answer was: “No, not at all.” Boys and girls between the ages of 12 and 16 reacted by laughing and calling the examiner a clown. Children between the ages of 5 and 6 . . . later [said] “The doctor was very nice, but it's funny how bad-mannered he is; he thumbed his nose at all of us.” Children 2 to 4 years old sometimes took an object – a ball, for instance – and threw like the examiner [had thrown], but they were merely playing with him. (Lhermitte, Pillon, and Serdaru, 1986, p. 329)

In contrast, almost all patients with frontal-lobe damage began imitating the experimenter with his first gesture. They did this seemingly without surprise, and with careful attention to detail. For example, when pretending to light a candle, no imitating patient

forgot to “blow out the match.” Later, when interviewed about the experiment, all patients could remember all of the gestures. Strikingly, “After being told not to imitate, most patients [nonetheless continued to] display the same IB” (p. 329). (With this study, Lhermitte, Pillon, and Serdaru were describing a “new” variant of utilization behavior: “imitation behavior” (IB).) IB was found in 28 of 29 patients with frontal lobe lesions in the study's sample. The 29th, recently diagnosed with a right frontal glioma (a tumor originating in glial cells), began to spontaneously exhibit IB three weeks later. Overall intellectual functioning of these patients was numerically, but not statistically, lower than that of control subjects. (In the 1940s, Hebb had reported that patients with PFC lesions can have normal IQ scores.) The frontal patients had varied damage, with inferior frontal gyrus involved in the majority of the patients. In contrast, IB was only seen in one of 21 non-frontal lesion control patients (and the factors behind this one patient's performance were characterized as “doubtful”).

Lhermitte suggested the term “environmental dependency syndrome” for even more exotic behavior by patients that he observed outside of the laboratory. One, who had undergone a large left-hemisphere frontal lobectomy to remove a glioma, was invited to the doctor's apartment. Upon walking past the bedroom and seeing the bed, the patient undressed, climbed into bed, and prepared to sleep! Subsequently, other neurologists have adopted “environmental dependency syndrome” as a blanket term encompassing all these types of behavior. In summarizing them, two other behavioral neurologists, Knight and D'Esposito (2003), have noted that:

It has been postulated that the frontal lobes may promote distance from the environment and the parietal lobes foster approach toward one's environment. Therefore, the loss of frontal inhibition may result in overactivity of the parietal lobes. Without the PFC, our autonomy in our environment would not be possible. A given stimulus would automatically call up a predetermined response regardless of context. (p. 262)

Next, let's consider how inhibition has also been implicated in a classic laboratory finding in lateral frontal-lobe patients – perseveration on the Wisconsin Card-Sorting Task (WCST).

Perseveration

In the late 1950s, decades of research on the effects of various brain lesions on cognitive flexibility had yielded contradictory and equivocal results, with some studies implicating the frontal lobes, some posterior regions, and some one hemisphere over another. In 1963, this debate was resolved, and a new avenue for research opened up, by a single study. Brenda Milner (of course) had the sagacity to realize that tasks that had been used in previous studies had the dual drawbacks of lacking sensitivity and lacking clearly defined psychological underpinnings. She had the insight to select a task that was likely to yield clearer results (*Tip 14.2*). And she had the benefit of working at the Montreal Neurological Institute (MNI), where she could test neurosurgical patients pre- and post-operatively – a repeated-measures design, of course, affording much greater statistical power.

The WCST was designed by (University of Wisconsin-based) psychologist David Grant (1916–1977) and his graduate student Esther Berg, who modeled it in part on a task by the German psychologist Egon Weigl, and in part on problem solving tasks developed for monkeys, and shown to be sensitive to frontal-lobe damage, by Grant's colleague Harry Harlow (1905–1981; developer of the Wisconsin General Testing Apparatus (WGTA; *Chapter 5*). As illustrated in *Figure 14.1*, the WCST consists of a stack of cards for which subjects must first learn, and then apply, a sorting rule. Milner tested 53 patients pre- and post-op: 18 with excisions invading dorsolateral PFC, seven with excisions of orbitofrontal and temporal cortex, and the remainder with excisions of temporal, parietal, or temporal-parietal-occipital cortex. The average number of categories achieved decreased for the dorsolateral PFC group from pre-op (3.3) to post-op (1.4) whereas this measure was effectively unchanged

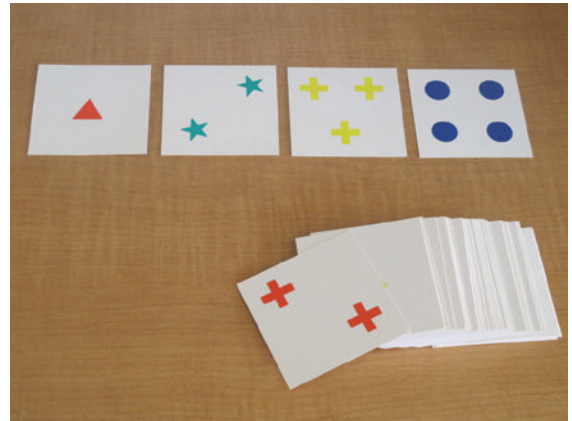


FIGURE 14.1 The Wisconsin Card-Sorting Test (WCST), as the subject sees it. She is given the stack of 128 response cards, and instructed to sort them by placing each, one at a time, under the appropriate “stimulus card” (“wherever you think it should go”), but with no further instruction. Each response card varies along three dimensions: color, shape, and number; each level of each dimension being represented across the four stimulus cards. In the example shown here, the two red crosses on the response card at the top of the stack match the *color* of the stimulus card on the left, the *number* of the next stimulus card to the right, and the *shape* of the next stimulus card to the right. The first sorting dimension is color, which the subject must learn by trial and error. The only feedback she receives throughout the task is whether each sort is “right” or “wrong.” After 10 correct consecutive trials (i.e., after she has learned to sort to color), the sorting dimension is switched (without warning) to shape. After 10 correct consecutive sorts to shape, it is switched to number. The procedure continues through another full cycle of the three sorting dimensions, or until all 128 cards have been placed. There are three dependent measures: number of categories achieved (i.e., number of dimension-changes + 1), number of perseverative errors, and number of nonperseverative errors. The former is a choice that would have been correct prior to the most recent switch. Source: Photograph by Michael Starrett, in the author's laboratory.

Tip 14.2

Decades later Brenda Milner told this author about writing to “the famous Dr. Grant,” who was “so gracious” in the return correspondence that it included the pack of WCST cards that she then used for the 1963 experiment.

for all other groups (overall mean, 4.6 to 4.7; for the orbitofrontal + temporal group, a slight decrease from 5.3 to 4.9). The culprit was **perseveration** (see caption to *Figure 14.1*; the term derives from Old French, and

before that Latin, for *persistence* and *stubbornness*). For dorsolateral PFC patients, perseverative errors increased, from 39.5 pre-op to 51.5 post-op, whereas the average for other groups declined from 20.0 to 12.2; for the orbitofrontal+temporal group they increased slightly from 9.3 to 12.0). Note that evaluation of pre-op performance needs to take into account that the surgeries for all patients targeted tissue that was epileptogenic, and so likely already contributing in only an impaired fashion, if at all, to task performance.

A perseverative error would seem to be one of a failure of inhibition, in that the patient “can’t stop himself” from repeating a choice already known to be incorrect. Indeed, as Milner (1963) notes, “perhaps the most interesting finding after frontal-lobe injury has been the dissociation between the subject’s actual responses to the stimulus-objects and his ability to verbalize what those responses should be.” Many patients were “able themselves to say ‘right’ or ‘wrong’ as they placed the cards, and yet did not modify their responses accordingly” (p. 99). That is, it was literally the case for some patients that their words indicated (well, suggested) that they knew what they were supposed to do, but that they couldn’t control their skeleto-motor responses accordingly.

It will be worth remembering, when we get to emotions and social behavior in later chapters, that the bias in the sample of Lhermitte, Pillon, and Serdaru (1986), who showed impairments on tasks that entailed direct social interaction, was toward damage in inferior regions of PFC, whereas in the Milner (1963) study with the WCST, it was much more a test of “cold reasoning,”

removal of the inferior and orbital portions of one frontal lobe causes no significant change, even if combined with a temporal lobectomy. One patient in seven showed a deficit in performance after such a removal, and this proportion does not differ from that showing loss after posterior excisions. In contrast to these negative results, all removals which encroached upon the superior frontal region were associated with poor performance on the sorting test. The critical lesion cannot be more precisely defined, although area 9 of Brodmann was implicated in most cases. (p. 97)

And so, the story thus far is that a defining function of the PFC, and of cognitive control, is the overriding

of prepotent responses (often summarized as “inhibition”), often triggered by environmental stimuli. Can we find physiological evidence for it?

Electrophysiology of the frontal-lobe syndrome

Robert Knight is a behavioral neurologist who got his scientific training with EEG electrophysiologists Steven Hillyard and Robert Galambos, senior authors on the auditory evoked-potential (AEP) studies that we considered in *Chapter 4*. Among Knight’s noteworthy contributions to cognitive neuroscience have been studies of ERPs in patients with PFC damage. *Methodology Box 14.1* illustrates how systematic deviations of ERP components from their typical pattern can provide insight about, in this case, mechanisms underlying cognitive control.

Integration?

The two main findings from the study described in *Methodology Box 14.1* highlight two distinct processes whose disruption may contribute to the lateral frontal-lobe syndrome: the filtering of irrelevant information, and the focusing of attention on behaviorally relevant information. Do either of these mechanisms help us understand the environmental dependency syndrome, or the phenomenon of perseveration? It’s not immediately apparent to me that they do. Whereas the Chao and Knight (1998) ERP findings relate to perceptual processing, the earlier findings that we reviewed related to the control of behavior. The environmental dependency syndrome is characterized by a remarkable influence of external stimuli and contexts on behavior, one that supersedes the autonomous control that is central to the human condition. Perseveration on the WCST occurred after a behavioral set was learned and repeated (reinforced), then needed to be changed when reward contingencies changed unexpectedly. Is it possible to unify these disparate findings under a single framework, or is cognitive control just a term that actually corresponds to a collection of task-specific proficiencies that don’t share any meaningful underlying relation, apart from a common address in the lateral PFC? This is precisely

METHODOLOGY BOX

14.1 Measuring ERPs in patients with focal cortical lesions

In 1995, Robert Knight and then-student Linda Chao reported that patients with lateral PFC lesions were disproportionately impaired on (5 second) delayed yes/no recognition of environmental sounds when distracting tone pips were played during the delay period. Going back to Malmö (1942), this is an oft-replicated consequence of lateral PFC damage. For a follow-up study, Chao and Knight (1998) repeated the experimental procedure, but while recording the EEG. To fully appreciate their study, it's worth taking a moment to flip back to *Figure 4.5* and refresh your memory about the organization of the AEP into early-, middle-, and long-latency components. What we're about to consider here is a masterful application of the logic, mentioned in that chapter, that "sensory evoked responses ... are stereotyped, stable responses that are effectively identical in all neurologically healthy individuals[, making] them very effective "tools" for studying the neural bases of many different behavioral and physiological states."

For the follow-up ERP study, Chao and Knight (1998) obtained ERPs during task performance from seven patients with unilateral damage to lateral PFC, and from matched control subjects. The previous behavioral finding of exaggerated sensitivity to distraction in the PFC group was replicated. To understand why, they examined the middle-latency components of the AEP (MAEP) to the distracting tone pips, because the MAEP is known to reflect the initial cortical processing of the auditory signal. From *Figure 14.2.A* one can see that several peaks in the MAEP are larger for patients than for control subjects. The proximal conclusion that can be drawn from this finding is that the patients' memory for the sample stimulus

FIGURE 14.2 Auditory evoked responses (AEP) in PFC-lesioned patients performing auditory delayed recognition. Source: Chao and Knight, 1998. Reproduced with permission of The MIT Press.

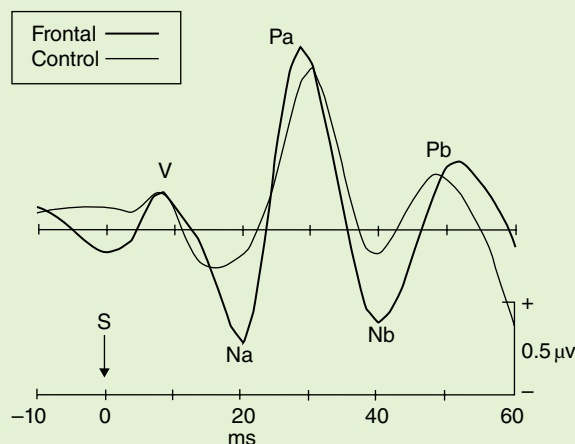


FIGURE 14.2.A The middle-latency components – MAEP – to tone pips played during the 5 s delay period of an auditory delayed-recognition task, for patients with lateral PFC lesions and control subjects.

Methodology Box 14.1 *Continued*

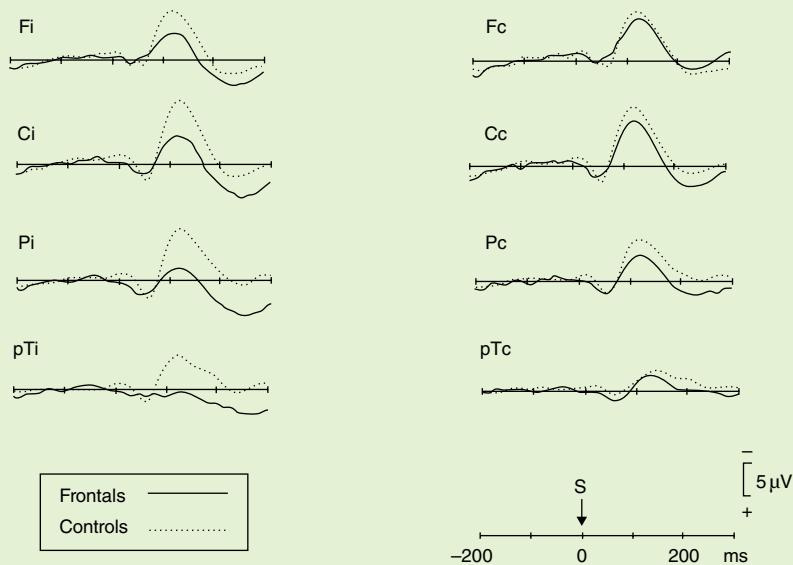


FIGURE 14.2.B A long-latency component of the AEP, the N1, measured at various locations on the scalp (“i” = ipsilateral; “c” = contralateral).

may have been disrupted by the abnormally high level of perceptual processing of the distracting stimuli. The inference about mechanism that follows from this is that one function of the PFC is to control the processing of incoming sensory stimuli, such that this processing doesn't interfere with behaviorally relevant internal representations. Further evidence for this interpretation comes from the fact that, in the patients, the magnitude of the Pa component of the MAEP predicted the magnitude of their behavioral deficit. The term often used for this hypothesized mechanism is “filtering,” as though the PFC imposes a filter that selectively passes only some incoming sensory information. Indeed, there's good evidence from the cat and from the monkey that PFC literally controls just such a filter via its projections to the thalamic reticular nucleus (TRN; see *Chapter 3* and *Chapter 4*). In the cat, these projections have been shown to regulate the magnitude of the AEP.

A second finding from this study was seen in a long-latency component of the AEP (LAEP) to sample stimuli, the N1. *Figure 14.2.B* shows that the N1 is smaller across the scalp in the PFC-lesioned patients. The auditory N1 is believed to arise from auditory association cortex, and previous research from Knight and colleagues has shown it to be disrupted in PFC patients in conjunction with impaired performance on tests of selective attention. Thus, Chao and Knight (1998) suggested that the reduced N1 in this study may have reflected an impairment in the ability of their patients “to focus attention on task-relevant stimuli” (p. 173) (*Tip 14.3*).

Tip 14.3

How do you think that the notion of “focusing attention on task-relevant stimuli” relates to that of encoding information into STM? What do you predict would have been the performance of the patients on a surprise subsequent memory test for sample items administered at the end of the ERP session?

the kind of question that computational modeling is good at addressing, and so it's to there that we now turn our attention. (*Web Clip 14.1* illustrates how the networks underlying cognitive control are being studied via fMRI scanning of frontal-lobe lesion patients.)

MODELS OF COGNITIVE CONTROL

Developmental cognitive neuroscience

The phenomenon of perseveration isn't restricted to behavioral neurology. Neurologically healthy adults will generate prepotent or habitual responses in inappropriate circumstances when their working memory is at capacity (such as in a dual-task paradigm), or when otherwise under stress. Perseveration is also a classic feature of cognitive development. The Swiss developmental psychologist Jean Piaget (1896–1980) codified this as the “A-not-B” error, seen in infants younger than 12 months of age.

The A-not-B error

To demonstrate the A-not-B error, a toy is placed under one of two boxes (box A) that are both within reach of an infant, and she displaces the box to retrieve the toy. This is repeated several times. On the critical trial, the infant watches while the toy is hidden under box B, but she nonetheless reaches for it under box A. Sometimes on the critical trial, in a behavior reminiscent of the dissociation between verbal report and card placement described by Milner (1963), the infant will look at box B (where the toy is hidden) yet reach for box A. Piaget ascribed this error to an immature

Tip 14.4

These findings provide an interesting example of how one's scholarly perspective can influence how one interprets data. Some have seen the PFC lesion findings as evidence that the theoretical construct of object permanence may not exist, and that the A-not-B error in young children is attributable to immature circuitry that supports a cognitive function like working memory, inhibition, or attention. Others have simply concluded that knowledge of object permanence must be supported by the PFC.

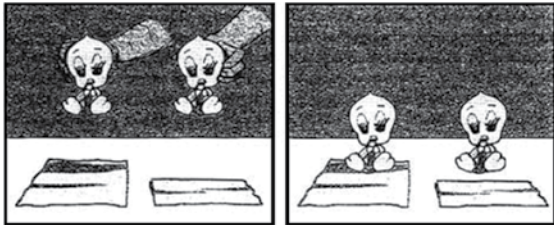
sense of “object permanence,” the understanding that objects continue to exist after they disappear from view. More recently, in 1989, the developmental cognitive neuroscientist Adele Diamond and Patricia Goldman-Rakic, with whom Diamond did a post-doctoral fellowship, documented the phenomenon in young rhesus macaques, replicated it qualitatively in adult monkeys with dorsolateral PFC lesions, and showed the absence of the effect in adult monkeys with PPC lesions. It is known that the primate PFC is among the last brain areas to mature, and that, in the human, it (along with other brain regions) undergoes rapid development during the first few years of life (*Tip 14.4*).

Is the A-not-B error a failure of working memory?

The association of the A-not-B error with the PFC led to proposals that succeeding on the task depends on working memory, and that underdeveloped working memory abilities underlie the error in infants. (Note that the Diamond and Goldman-Rakic findings were published at the same time as the latter's neurophysiological recordings from PFC.) The working memory view was soon challenged, however, by studies with human infants using the procedure illustrated in *Figure 14.3.A*, in which all the information needed to solve the task was always visible to the infant (i.e., no overt demands on working memory were made), yet it still produced the A-not-B error. What *Figure 14.3.B* illustrates is a schematic of a neural network simulation by developmental cognitive

FIGURE 14.3 Modeling perseveration in the A-not-B error. Source: Stedron, Sahni, and Munakata, 2005. Reproduced with permission of The MIT Press.

A trial



B trial

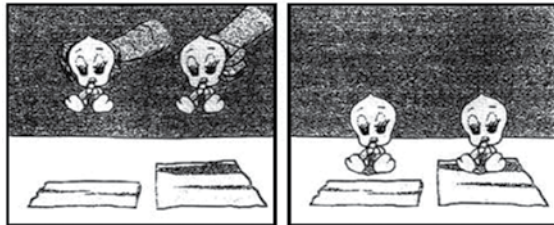


FIGURE 14.3.A An example of an experimental procedure that produces the A-not-B error in infants, even though there is no overt working memory component. Infants sit on a parent's lap, and watch as two toys are placed side-by-side, out of reach, but one on a piece of cloth and one not on a piece of cloth. (The parent is occupied so that s/he won't bias the child's performance.) The closer ends of the cloths are within reach, and the infant knows to pull on the cloth on which the toy sits, in order to retrieve it. This is repeated several times. On the B trial, the cloths are switched, such that it is now the cloth on the right that should be pulled to retrieve the toy. Infants under the age of 12 months, however, commit the A-not-B error.

neuroscientist Yuko Munakata and students. In offering an explanation for the A-not-B error, it addresses several other questions that have arisen in this chapter: it provides an explicit account for why perseveration happens, and what might be the PFC mechanism that controls it; it proposes a mechanism that would unify the ERP findings from Chao and Knight (1998) with the observations of perseveration and of environmental dependency syndrome; and it offers one answer to the question of whether the lateral frontal-lobe syndrome is fundamentally one of a loss of inhibitory control. And it's all in *Research Spotlight 14.1*.

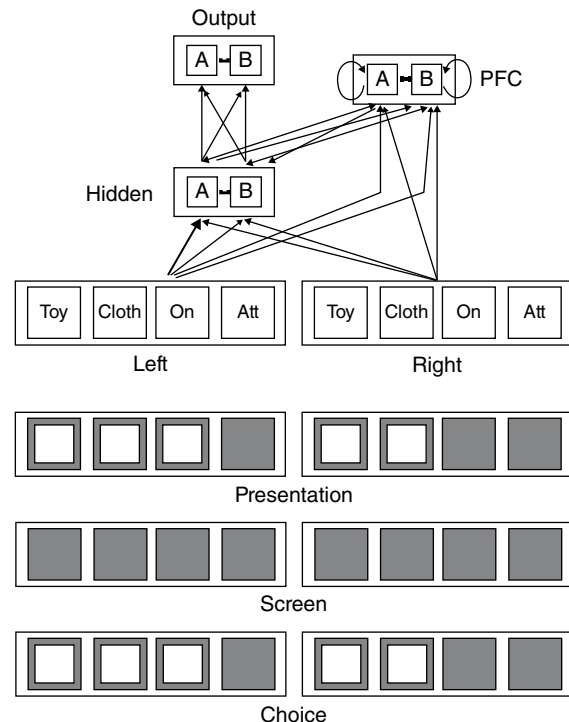


FIGURE 14.3.B Architecture of the model (top), and, below, strength of inputs to the model at three epochs of the trial. (A screen is used for a brief period of time in behavioral studies, to prevent the infant from reaching while the stimuli are still being positioned.) Although the authors do not state this explicitly, the input layer can be construed as extrastriate visual areas, the hidden layer as the parietal reach area (*Chapter 6* and *Chapter 7*), and the output layer as motor cortex. ("Att" can be disregarded; it was used for a variant of the task in which the toy is attached to the cloth on which it sits.)

Generalizing beyond development

Implied in the Stedron et al. (2005) results (*Figure 14.3*; *Research Spotlight 14.1*) is something that was also implied in my chapter-opening elevator anecdote, and stated explicitly by Miller and Cohen (2001): lurking within all of us is a machinery whose default mode of operation is to respond to the most salient stimulus in the environment, or act in the most habitual, stereotyped way. Explicit in their model is the idea that cognitive control may not be implemented

RESEARCH SPOTLIGHT

14.1 Overcoming perseveration in the A-not-B task

The simulation of Stedron et al. (2005) used a neural network simulation environment created by O'Reilly (of McClelland, McNaughton, and O'Reilly renown) and Munakata (2000) that is grounded in the principles of the PDP framework, and that explicitly simulates the properties of biological networks by modeling membrane potential within each unit and determining the propagation of activity from one layer to another via an equation that converts the membrane potential into an activation value (a proxy for the rate and number of action potentials). Another important feature of the model is the Hebbian strengthening of connections, which occurs when anatomically connected units are simultaneously active.

Figure 14.3.B illustrates an A trial in the A-not-B task, for which the attributes of *Toy*, *Cloth*, and *On* are activated for the left side (i.e., the A side) of the display, and the attributes of *Toy* and *Cloth* are activated for the right (i.e., B) side. It is clear from this configuration of activation values that units representing choice A in the hidden layer and in the PFC layer receive more activation than do those representing choice B, because $3 > 2$. Thus, choice A receives more activation in the output layer, and the correct choice is made. As this process is repeated across several trials, the connections between the left-sided input layer and the representation of A in the hidden layer, and between the representation of A in hidden layer and the representation of A in the output layer, get strengthened by the Hebbian learning rule. This models, in a very concrete way, prepotency. A consequence of this prepotency is that, on the critical B trial, even though $2 < 3$, the strength of the drive received by the A unit in the Output layer is greater than that received by the B unit in the Output layer. This is because multiplying the input value of 2 by larger numbers (due to stronger connections) can yield a larger result than does multiplying a starting value of 3 by smaller numbers. End result? The model will choose Output A – it will perseverate. How can this problem be overcome? By top-down input from the PFC.

Unlike with the classic A-not-B task, where the toy is hidden from view under a box, success on this always-visible variant comes earlier in development: 9-month-olds perseverate with this procedure, but 11-month-olds do not. Stedron and colleagues simulated the maturation of PFC from 9 to 11 months by progressively strengthening recurrent weights between units within the PFC layer, thereby allowing it to progressively enhance its ability to sustain a representation of the input side that had a higher value, regardless of the strength of learned connections from Input to Hidden to Output layers. Their finding was that at a certain level of PFC “maturity,” the ability of this layer to sustain a representation of the correct response provided a sufficiently strong top-down boost of the hidden layer that the B output was selected, despite the structural advantage of the representation of A. (*Web Clip 14.2* describes a translation of this principle into a behavioral intervention for impulsive children.)

by inhibiting inappropriate responses but, rather, by selecting the most appropriate response and activating it to the point that it will win out over other alternatives. This is equivalent to the idea of biased competition (*Chapter 6*). Can the ERP findings of Chao and Knight (1998; *Methodology Box 14.1*) be understood

in this framework? Recall that, at a descriptive level, they suggested impairments of “filtering” and of “focusing.” The former, that PFC-lesioned patients showed abnormalities in the MAEP and LAEP, can be explained as dysfunction at the *site* of cognitive (or attentional) control – a failure of sensory filtering at

the level of the thalamus due to a weakened top-down signal to the TRN. The latter, manifesting as a diminished N1, could be the result of a weakened top-down selection signal. If one were striving for a unifying account of PFC function, one might posit that these could be the same signals that, when directed to the appropriate networks in parietal and/or motor cortex, would select the appropriate response over the prepotent but inappropriate one.

But can the “unitary mechanism” simulated in the Stedron et al. (2005) model really account for the lateral frontal-lobe syndrome in the adult? There are at least a couple of arguments that one might raise. First, the A-not-B error is a behavior that infants grow out of before they can even walk. Second, that simulation is limited to a drastically impoverished and constrained environment (only 3×2 inputs; only two choices of response). How applicable can it really be to understanding the richness and complexity of adult cognition and neurological syndromes? There’s an additional concern that may or may not have been raised in your mind by the Stedron et al. (2005) model, but that has certainly been on mine since we considered the Hamker (2005) model back in *Chapter 8*, and that’s *Who is controlling the controller?* For all of the computational models that we’ve considered, “who” is keeping track of the rules governing the task that’s being performed? For a task like the WCST, how does simply receiving the feedback “wrong,” as every subject does each time the sorting dimension is surreptitiously changed, prompt us to try a different strategy? The facetious answer is that there’s that “little man” inside our head, the homunculus, who knows what different parts of the brain are supposed to do in different situations, and tells them how, and when, to do it. Now of course, it’s been hundreds of years since any serious scientist or philosopher may have proposed such a thing. On the other hand, the majority of contemporary models that seek to understand one aspect of cognition depend, in effect, on what Rougier, Noelle, Braver, Cohen, and O’Reilly (2005) refer to as “some unexplained source of intelligence” (p. 7338) to do the things that the model can’t. The model that we’ll consider next develops the capacity for error correction and cognitive control without resorting to a homunculus, or leaving this question unspecified.

What makes the PFC special?

I said earlier that I didn’t want to endorse a “PFC = cognitive control” perspective, and yet we’re more than halfway through the chapter and what have we discussed? (Exclusively!?) Consideration of this neural network simulation by Rougier et al. (2005) will accomplish two goals in this regard. First, it will consider some aspects of the PFC that justify this bias, by calling explicit attention to unique properties of this region and the neurons that make it up, relative to other parts of the brain. Second, it explicitly considers how the PFC must interact with other brain systems in order to accomplish cognitive control. (I.e., it can’t literally do it all by itself.)

Properties of the PFC

The Rougier et al. (2005) model of PFC includes three properties that distinguish it from other areas of cortex:

- Sustained activity that is robust against potentially interfering signals: PFC pyramidal neurons have the property of bistability – the ability to operate in an “up” state or a “down” state. This is partly due to the density and distribution of D1-type dopamine receptors along their dendrites, and so the state that they occupy at any point in time is influenced by dopamine levels. When in the up state, a high density of NMDA receptors and recurrent connections with other pyramidal cells can support sustained, elevated levels of activity, such as those illustrated in *Figure 13.1.A*.
- Capacity for rapid updating of activity patterns by control of bistability, i.e., switching from an active maintenance mode, when one pattern of activity is sustained so as to prioritize one type of behavior, into an “updating mode,” when the previous pattern is destabilized and a new pattern of activity can be established. This is effected by dopamine signals and by input from the basal ganglia, and turns out to be key for circumventing the homunculus problem, in that it helps “tell the PFC what to do.”
- Capacity for PFC modulation of other cortical areas (as we have already seen in this chapter with Chao and Knight (1998), and in *Chapter 8* with FEF implementation of the endogenous control of attention).

Before we can fully understand the model, however, there's a new concept that we need to introduce in *Thought Box 14.1*: the reward prediction error signal.

Influence of the DA reward signal on the functions of PFC

Turning now to *Figure 14.5*, we can see that this model is, for the most part, just a more-finely-carved-up variant of the model from *Research Spotlight 14.1*.

The finer grain of detail in the input layer, for example, will enable the simulation of tasks like the WCST, in which there are multiple stimulus dimensions, and multiple levels within each dimension. For our purposes, the key qualitative difference is the little red box in the upper right, labeled “AG,” for adaptive gating. This carries dopaminergic reward signals.

Not unlike PDP models that we've considered previously, the Rougier et al. (2005) model started from scratch, knowing nothing about its world. It was

THOUGHT BOX

14.1 The reward prediction error signal

Recall from *Chapter 7* that the neurotransmitter dopamine (DA) is implicated in motor control, with degeneration of the substantia nigra pars compacta being the proximal cause of Parkinson's disease. Because there were no data directly linking activity in these neurons to movement, however, Mexican neurophysiologist Ranulfo Romo and his Swiss postdoctoral mentor Wolfram Schultz set out to study this by recording from midbrain DA neurons in the monkey. To their surprise, they failed to find evidence that these cells responded to movement; instead, they seemed to respond to motivational and reward factors. Romo later noted “It took us about 5 years to publish the paper because nobody was going to buy this idea that those neurons were not associated with movement, but reward” (Downey, 2006, p. 14264).

Once they did get it published, it was quickly realized that these findings had profound implications for such topics as learning, motivated behavior, decision making, and addiction. For our present purposes, one important factor in controlling behavior is to be able to predict future events – the failure of an expected outcome to occur can be an important cue to change one's behavior. *Figure 14.4* summarizes some key properties of midbrain DA neurons that Schultz and colleagues have characterized, including the pause in firing that is observed when an expected (i.e., predicted) reward is not delivered. This is an error signal in that it indicates a mismatch between the prediction and what actually occurred.

Properties such those illustrated in *Figure 14.4* indicate several important facts about these neurons. One is that they receive information about sensory stimuli, such as the lights and tones that are used as conditioned stimuli, as well as about appetitive stimuli (juice, food) used as unconditioned stimuli. Another is that the drive from these inputs on midbrain DA neurons is modifiable with learning. A third is that some of these DA neurons also receive signals that convey the expected time of reward delivery. In addition to these empirical facts, midbrain DA neurons have two other properties that make them well suited for a key role in building and modifying predictions. First, they have the capacity to influence plasticity (and, therefore, learning). Second, they broadcast their signal diffusely, so that stimuli from all modalities can be used to make and evaluate predictions.

Before returning to neural network models of cognitive control, let's get a foreshadowing of what's to come in *Chapter 15* by concluding this thought box with a snippet of an admonition from Schultz about how to think about reward:

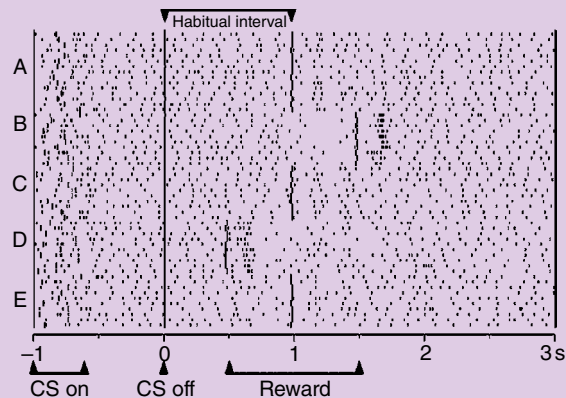
Thought Box 14.1 *Continued*

FIGURE 14.4 The reward prediction error signal. Each row shows activity in a midbrain DA neuron during a well-learned classical conditioning task in which the offset of the conditioned stimulus (CS) is followed by a reward one second later (the habitual interval). When the expected outcome occurs (groups A, C, and E) there is no change in firing rate. When the reward is unexpectedly delayed (B), there is a brief pause in firing – the reward prediction error signal – then an elevated burst upon the unexpected, later delivery. When the reward is unexpectedly delivered early (D) there is an elevated burst signaling its delivery, but no suppression at the normal reward delivery time. This type of reward signaling is generally more prevalent in the neurons in nuclei A10 and A8 (i.e., the ventral tegmental area, VTA) that project to cortex than in those of nucleus A9 (the substantia nigra pars compacta) that project to the neostriatum. Source: Schultz, 2006. Reproduced with permission of Annual Reviews.

Primary sensory systems have dedicated physical and chemical receptors that translate environmental energy and information into neural language. Thus, [their] functions ... are governed by the laws of mechanics, optics, acoustics, and receptor binding. By contrast, there are no dedicated receptors for reward ... [and so] the investigation of reward functions requires behavioral theories that can conceptualize the different effects of rewards on behaviors ... Animal learning theory and microeconomics are two prominent examples of such behavioral theories. (Schultz, 2006, p. 91)

Reward, uncertainty, contingency, probability, expected value; these don't physically exist in the world, but must be constructed and represented neurally, in what Weigl (1941) called *Abstraktionsprozesse*.

trained by learning to discriminate and make decisions about stimuli via the four training tasks (see legend to *Figure 14.5.B*). A critical fact about this process, however, is that instead of the teaching/learning process being implemented “*deus ex machina*,” by a scientist delivering a backpropagated error signal each time she observed that the model got an answer wrong, it was implemented by the DA-mediated

reward mechanism in a manner that emulated what we know about how this system works. Think back to the example of Pavlovian conditioning that we considered in *Chapter 10*. What Schultz and colleagues did in some of their studies (as summarized in *Thought Box 14.1*) was literally to monitor the responses of midbrain DA neurons throughout the entire process of learning a CS-US pairing. As a result,

FIGURE 14.5 Prefrontal cortex and flexible cognitive control: rule-guided behavior without a homunculus.
 Source: Rougier, Noelle, Braver, Cohen, and O'Reilly, 2005. Reproduced with permission of PNAS Group.

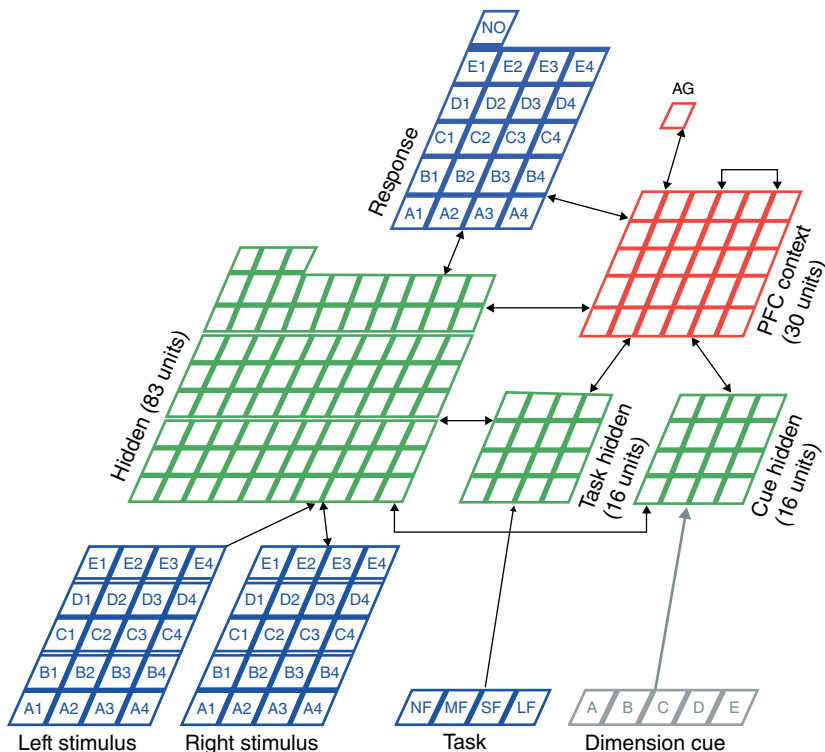


FIGURE 14.5.A The architecture of the model. “AG” is the “adaptive gating” unit, which corresponds to the ascending dopaminergic reward-related signal.

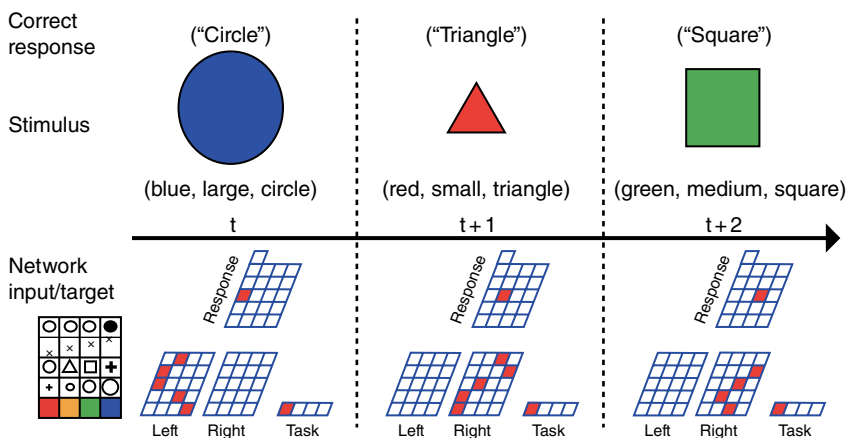


FIGURE 14.5.B Illustration of how high-dimensional inputs were fed into the model. The key in the lower left illustrates how one stimulus dimension is represented in each of the five rows of the input array – color, size, shape, location, and texture – and the columns correspond to different levels of each dimension. From left to right is an illustration of how the presentation of three consecutive stimuli (top row) would be represented and processed by the network. The “task” layer tells the model which of four tasks to perform: naming the feature in a particular stimulus dimension; determining whether two features match according to a particular stimulus dimension; indicating which item is larger; and indicating which item is smaller.

we know, for example, that in pairings where the CS precedes the US in time (as is illustrated in *Figure 14.4*), there is initially a phasic increase in DA associated with reward delivery. (This has been interpreted by some models as signaling that “the current state is better than expected” (Montague, Dayan, and Sejnowski, 1996, p. 1944).) Over time, as the pairing is learned, the US ceases to elicit this burst of DA. (By definition, once the pairing has been learned, the arrival of the US is now expected. Thus, its delivery can no longer be characterized as being “better than expected.”)

The architecture of this network incorporated the fact that PFC is much more densely innervated by DA fibers than are posterior cortical regions, and a consequence of this was that the hidden layer, corresponding to temporal and parietal cortex, developed differently than did the PFC layer. The hidden layer developed to represent stimulus features, and thus the ability to discriminate between features. The PFC, in contrast, developed representations of the rules of the tasks that the network was trained on: how to select the correct name when presented with two objects that both have names; or how to decide if a medium-small-sized, yellow, vertically striped square located in the lower-left of the screen is the same as or different from a small, green, vertically striped square located in the upper-right if the matching dimension is size. This PFC-based representation of abstract rules (there it is again, the *Abstraktionsprozesse!*) facilitated task performance by providing top-down excitation to the representations of the relevant stimulus dimension in the hidden layer of the network. We have already seen that this is one of the functions of the PFC. The new insight for us is the mechanism by which it influenced the network’s performance on the WCST.

Combining a reward prediction error signal with a flexible PFC to overcome perseveration

At the end of the training process, when the model was mature, Rougier et al. (2005) introduced the WCST. The task was simulated by the feature-naming task, but minus any input to the “task” layer or to the “dimension cue” layer that would indicate which dimension was to be named. Thus, it had to rely on trial-and-error learning, when presented with a medium-small, yellow, vertically striped square located

Tip 14.5

The top-down bias of the PFC on the hidden layer, favoring one stimulus dimension over others, is a simulation of the “selection negativity” component of the ERP that is illustrated in *Figure 6.17B*.

in the lower-left of the screen, to decide whether to give the response “medium-small,” “yellow,” “vertical stripes,” “square,” or “lower-left.” Feedback was conveyed by the DA reward signal, operationalized by the AG unit, following the dynamics described above. When, in the course of trial-and-error guessing, the network stumbled upon the correct dimension by naming the color of a stimulus, the PFC received a phasic burst of DA. And here’s where the dopaminergic control of PFC bistability comes in. The response of the model’s PFC to this burst of DA was to switch into its up state, in effect “locking in” its active representation of the dimension *color*. (There is considerable biophysical, neuropharmacological, and neurophysiological evidence that a phasic increase of DA can trigger such a state change in PFC pyramidal neurons.) The sustained representation of the category of *color* in the PFC, in turn, biased the competition within the hidden layer of the model in favor of representing color relative to other stimulus dimensions (*Tip 14.5*). This state of affairs persisted across the next several consecutive trials. On the switch trial, when (unknownst to the network) the sorting rule was changed to *shape*, the network blithely named the color of the stimulus, just as it had in the previous several trials. This time, however, it was given the feedback “incorrect.” The failure to receive the predicted reward produced a reward prediction error signal, which had the effect (again, based on experimental evidence) of shunting the PFC out of its active maintenance mode and into its updating mode. This destabilized the PFC activity pattern corresponding to the dimension of *color*, thereby permitting a different activity pattern, corresponding to a different stimulus dimension, to emerge.

One take-home from all this is that the role of the PFC in enabling flexible adaptation to unexpected changes in the environment isn’t mysterious; isn’t the result of “some unexplained source of intelligence.”

Tip 14.6

The details of how midbrain DA neurons compute information related to reward factors are not yet well understood. A particularly intriguing question is the source, and computation, of information representing the expected time of reward delivery.

Rather, it results from a combination of the unique biophysical and connectivity properties of its neurons, its uniquely dense innervation by midbrain DA neurons, and the reward signaling properties of these DA neurons (*Tip 14.6*).

NEURAL ACTIVITY RELATING TO COGNITIVE CONTROL

Having considered what can happen when control goes awry, and some computational principles underlying its implementation, we'll conclude this chapter by considering some neurophysiological correlates of cognitive control. This is a fascinating area of research, and one in which, as we've seen elsewhere, new findings can sometimes raise more questions than they answer.

Error monitoring

The ERN/Ne

In the early 1990s, two research groups, one in Dortmund, Germany, and one in Ann Arbor, Michigan, reported an ERP component that corresponded to making an error on speeded reaction-time tasks. The Dortmund group, led by Michael Falkenstein, called it the Ne (error negativity) and the Ann Arbor group, led by William Gehring, called it the ERN (error-related negativity). The fact that it started out with, and, to some extent, still goes by, two names, is emblematic of the enduring lack of consensus that there has been for what process(es) it is likely to index. Here, we will call it by its more common name, the ERN. As illustrated in *Figure 14.6*, the onset of the ERN occurs shortly before, or at, the moment of the erroneous button press, and peaks roughly 100 ms later. The reason for the

intense interest in the ERN is clear, as summarized by Gehring, Liu, Orr, and Carp (2012):

Cognitive control functions include processes that detect when control is needed – as when performance breaks down – and processes that implement control through changes in attentional focus and other strategic adjustments. Because an error is a salient marker that performance has broken down, the ERN is generally thought to reflect a process involved in evaluating the need for, or in implementing, control. (p. 233)

Subsequent research indicates that an ERN can be generated in at least two circumstances. One, illustrated in *Figure 14.6*, is the “response ERN.” The second is the “feedback ERN,” which peaks 250–300 ms after the delivery of negative feedback. Some consider these to be two manifestations of activity in the same error-processing system. The response ERN, however, differs from the feedback ERN in that its timing raises the fascinating question about whether it is accompanied by awareness. We've all had the experience of making a mistake, doing something wrong, and not realizing that it was a mistake until minutes, hours, days later. For example, “I thought that this new hairstyle was a good idea, until the barber finished and I saw the result in the mirror for the first time.” Similarly, in the WCST, one can't know on a switch trial that one has made an error until being told that the previous action was “wrong.” Here, in contrast, the question is whether, on a moment to moment basis, there are circumstances in which our nervous system knows that we are making a mistake even while the action is being carried out, tens if not hundreds of milliseconds before we will become aware of the error. One example is what just happened to me: in the previous sentence I typed “erro,” but didn't realize it until I saw the jumble of letters on my computer screen. This is a fascinating literature, full of nuance, that is reviewed in the referenced *Further Reading* by Wessel (2012). (Just did it again! I typed “Wellel”...)

Performance monitoring and the dorsomedial frontal cortex

Considerable evidence, from source localization of EEG and MEG data, from recordings from nonhuman primates, and from fMRI scans of similar tasks, indicate that the ERN is likely generated in frontal midline areas generally referred to as the dorsomedial

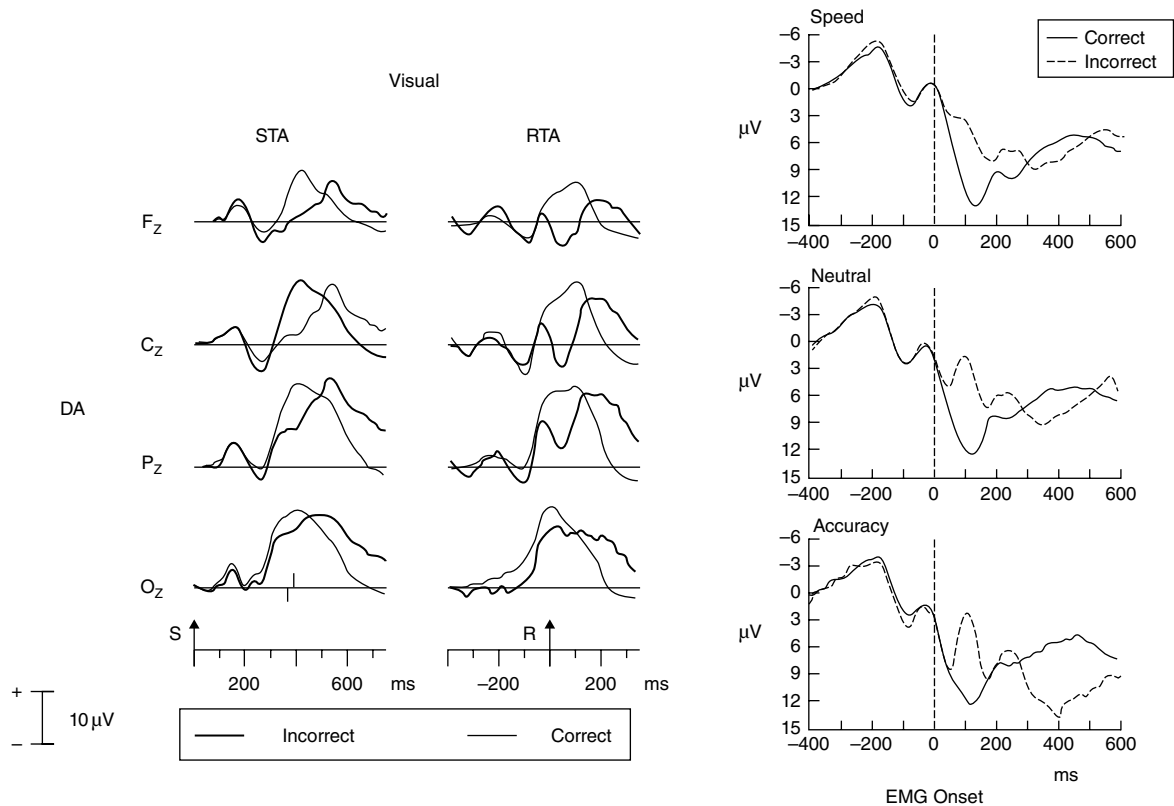


FIGURE 14.6 Two exemplars of the ERN. On the left, Gehring (2012) based on data from Falkenstein, Hohnsbein, Hoormann, and Blanke (1991), are ERP waveforms aligned to stimulus onset (STA; marked by the arrow on the timeline) or to the response (RTA) in a visual discrimination task. Note that the negative-going component on incorrect trials is only detectable for the latter. (Note, too, that negative is plotted down.) On the right, Gehring (2012) based on data from Gehring, Goss, Coles, Meyer, and Donchin (1993), are ERPs from three conditions of a flanker task, time-locked to the onset of the electromyogram (EMG) of the muscles controlling the finger. (In these plots, negative is plotted up.) Source: Gehring, Liu, Orr, and Carp, 2012. Reproduced with permission of Oxford University Press.

frontal cortex (dmFC), a term that encompasses the posterior anterior cingulate cortex (ACC) and the pre-supplementary motor area. (Note that, in this context, the area in question is much more extensive than the “DMFC” whose oculomotor functions were illustrated in *Figure 8.3*.) Indeed, as illustrated in *Figure 14.7*, and reviewed in *Thought Box 14.2*, this region has been implicated in many types of performance monitoring situations.

Oscillatory synchrony

A final problem of cognitive control that we’ll address here is one that was articulated by Earl Miller and his former student Timothy Buschman

in a 2013 review: “How are some thoughts favored over others?” (p. 216). The question is so fundamental to cognitive control that it’s almost tautological. It underlies, for example, how a behavioral response is selected when one is confronted with a neurologist making an insulting gesture, or when informed that the previously correct card sorting strategy is no longer correct. It underlies how certain stimuli are selected as to-be-remembered on a test of STM and others are ignored. There is now accumulating evidence that the attentional prioritization of information is accomplished by increasing the oscillatory synchrony between areas involved in processing that information. In particular, studies employing simultaneous neuronal recordings in

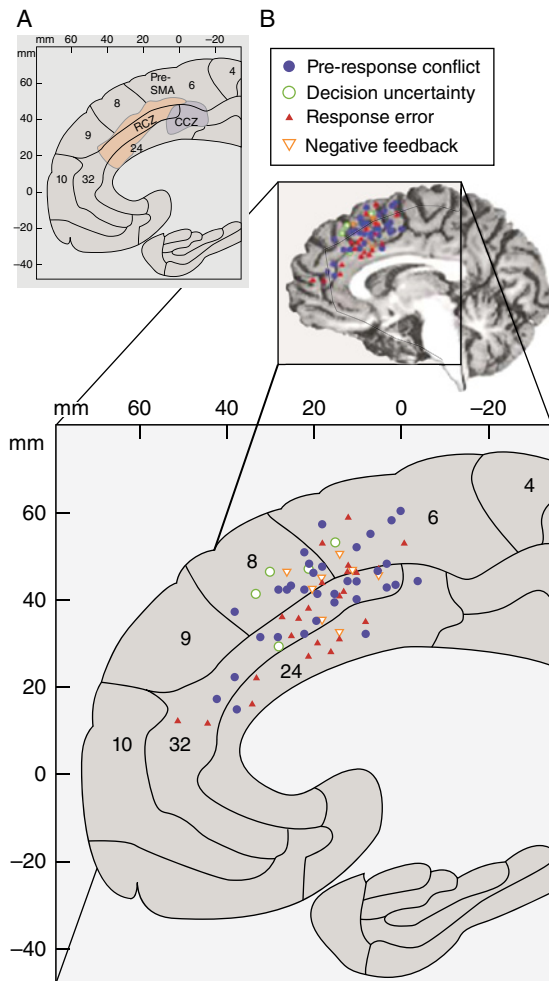


FIGURE 14.7 Performance monitoring and the dorsomedial frontal cortex. **A** illustrates a functional distinction proposed by Picard and Strick (1996), and **B** a meta-analysis from Ridderinkhof and colleagues (2004) that is summarized in *Thought Box 14.2*. Source: Ridderinkhof, Ullsperger, Crone, and Nieuwenhuis, 2004. Reproduced with permission of AAAS.

anatomically distal regions have revealed behaviorally relevant increases in synchrony between, for example, the FEF and V4, between V1 and V4, and between PPC and MT. This is a concept first introduced in *Chapter 2*, and illustrated in *Figure 2.11*. A question that remains, however, is the mechanism whereby anatomically distal regions are brought into synchrony. One candidate system that may

implement this mechanism for visual processing is the pulvinar nucleus.

Thalamic control of cortical synchrony

Figures 14.8 and 14.9 illustrate a study by Australian-born neurophysiologist Yuri Saalmann and German-born cognitive neuroscientist Sabine Kastner and colleagues, performed in Kastner's lab at Princeton University, that demonstrated a role for the pulvinar nucleus in synchronizing oscillations between V4 and TEO during a selective attention task. First, they acquired diffusion tensor images (DTI; *Methodology Box 2.1*) so as to find common projection zones from pulvinar to the two cortical regions. Next, they simultaneously recorded from overlapping receptive fields within the three regions simultaneously. The findings, illustrated in *Figure 14.9*, suggest that the pulvinar imposes synchronous rhythms on V4 and TEO during the delay separating the spatial cue and the stimulus array. This effect was strongest in the alpha band, and also statistically significant, though numerically weaker, in the gamma band.

Note that an important part of the evidence presented in *Figure 14.9* is the spike-field coherence data. Plots A, B, and C, in that figure show coherent LFPs in each of the pairs of regions depicted, and this is, indeed, the state that is believed to be important in order for, in this case, V4 and TEO to efficiently process information about the target stimulus. Further, a statistical analysis called a conditional Granger causality analysis (*Tip 14.7*) suggested that LFP oscillations in the pulvinar were “Granger causal” on those in the two cortical areas. That is, that the LFPs in the pulvinar seemed to

Tip 14.7

Granger causality analysis tests whether, say, past values of the pulvinar LFP help predict the current value of the V4 LFP, above and beyond the prediction derived from past values of V4 itself. In the context of the three-node model being considered here, *conditional* Granger causality of pulvinar on V4 also would be *above and beyond the prediction derived from past values of V4, itself, and of past values of TEO*.

THOUGHT BOX

14.2 One phenomenon, so many theories

Well, it started out as one phenomenon – the ERN/Ne. The flurry of research activity that it has engendered, however, has led to a much more complicated state of affairs. Traditionally, the dmFC has been thought of as part of the motor system, as exemplified by the organizational scheme of Picard and Strick (1996), illustrated in *Figure 14.7.A*. It distinguishes between a rostral cingulate zone (RCZ) activated in relation to complex motor tasks, and a smaller caudal cingulate zone (CCZ) activated during simpler tasks. The near-complete overlap between these two regions and the sites of activity identified by the many studies contributing to panel B in *Figure 14.7* illustrates very concretely the idea that we have contemplated elsewhere, that many aspects of high-level cognition can be construed as elaborations on more basic principles of perception and action.

Panel B of *Figure 14.7* shows the results of a meta-analysis, led by the Dutch cognitive neuroscientist Richard Ridderinkhof, of the foci of activity in neuroimaging studies that were designed to isolate one of four types of performance monitoring phenomena:

Pre-response conflict

Response conflict occurs when a task concurrently activates more than one response tendency, a classic example being the Stroop task, in which it's more difficult to name the color of ink in which a word is printed if the two aren't congruent. (I.e., there's more response conflict, indexed by longer reaction times, more errors, and greater activity in dmFC, when red ink is used to spell out the word "BLUE" than when it is used to spell out the word "RED" (a congruent stimulus) or the word "BOOK" (a neutral stimulus).) Response conflict processing can theoretically also occur after an incorrect response, as a residual representation of the correct response conflicts with the activated representation of the executed response. The canonical statement of the response conflict theory is presented in Botvinick, Braver, Barch, Carter, and Cohen (2001).

Decision uncertainty

Also referred to as "underdetermined responding," this is operationalized by conditions requiring a choice from a set of responses, none of which is more compelling than the others

Response error

This category includes the response ERN and the commission of errors on two-alternative forced-choice tasks. It is often associated with premature responses. One theory motivated by these effects draws explicitly on the dopamine reward-related signals summarized in *Figure 14.4*. Specifically, it posits that the RCZ processes these signals in accordance with the principles of **reinforcement learning** to improve task performance. A place to learn more about this perspective is Holroyd and Coles (2002).

Negative feedback

The canonical event for this category from *Figure 14.7.B* is one that would produce a feedback ERN. Included in this category are trials in economic games when an unexpectedly low monetary reward is given, or when a more abstract form of negative feedback is given (like being told that a response on the WCST is "wrong"). One idea for what unifies these types of trials is that they entail a comparison of expected vs. actual outcomes. Ridderinkhof, Ullsperger, Crone, and Nieuwenhuis (2004) also note that "Similar parts of the pmFC are activated by primary reinforcers such as pain affect and pleasant tastes, suggesting that the pmFC plays a general role in coding the motivational value of external events."

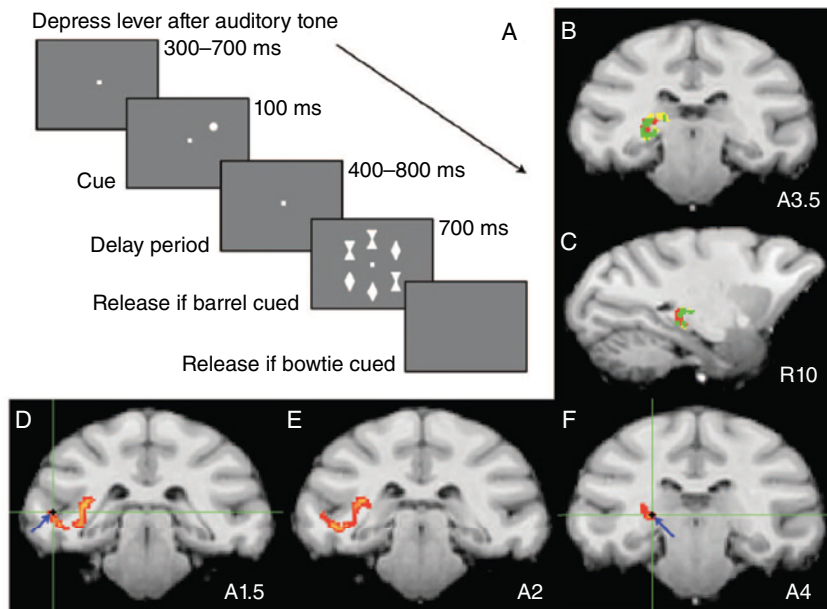


FIGURE 14.8 Thalamic control of interareal synchrony in cortex; the methods. The visual discrimination task of Saalmann, Pinsk, Wang, Li, and Kastner (2012) cued the monkey where to direct its attention, and its task was to release a bar during the display of the array if the stimulus occupying the cued location was a “barrel” [they look like diamonds to me] and to withhold its response until after array offset if the stimulus occupying the cued location was a “bowtie.” Panels **B–F** show the results of DTI analyses, with **B** and **C** showing voxels in the pulvinar nucleus connected with V4 (yellow), TEO (red) and both (green). **D–F** show, in sequential coronal slices, a fiber pathway between a “green” voxel in the pulvinar nucleus (in panel **F**) and TEO (in panel **D**). In **D** and **F**, a blue arrow indicates the location of a recording electrode. Source: Saalmann, Pinsk, Wang, Li, and Kastner, 2012. Reproduced with permission of AAAS.

lead and those in the cortical areas to follow. One must keep in mind, however, that an implication of the “local” part of “local field potential” is that none of these regions could “know” the dynamics of the LFP in the others – there is no physical way that the pulvinar LFP, itself, could drive LFPs in these cortical areas. This is why the spike–field coherence data are so important. What they show is a coherence between action potentials fired by the pulvinar and the LFP in V4 and TEO. It is through the temporal patterning of action potentials generated by an area that it can influence the LFP in its efferent targets. This may be just one specific case of a broader principle underlying the implementation of cognitive control.

The final conclusion from these data is that *Figure 14.8* and *Figure 14.9* illustrate a mechanism

whereby the pulvinar nucleus can establish a prioritized channel along which behaviorally relevant visual information will be processed. (These ideas and methods are considered in further detail in *Web Clip 14.3*.)

WHERE IS THE CONTROLLER?

In this chapter we have seen that the dorsolateral PFC plays a key role in the control of cognition. Damage to this structure leads to the dramatic and debilitating behavioral disinhibition characteristic of the lateral frontal-lobe syndrome. On closer inspection, we saw that the implementation of cognitive control involves many stages, and many “players.” Midbrain dopamine neurons send reward-related signals that instruct the

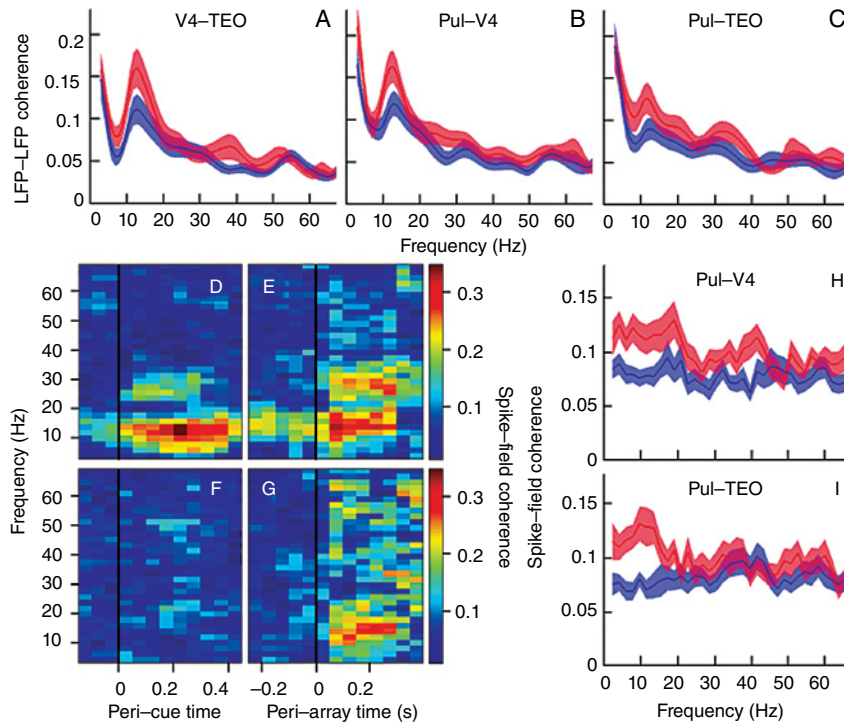


FIGURE 14.9 Thalamocortical control of interareal synchrony in cortex; the results. **A–C** show coherence of LFPs (ordinate) at frequencies from 0–60 Hz (abscissa) between **(A)** V4 and TEO, **(B)** V4 and pulvinar, and **(C)** TEO and pulvinar; with red plot corresponding to trials when the attention was allocated inside the receptive fields of the regions and blue when it was allocated elsewhere. **D–G** show the “spike–field coherence” between pulvinar and V4 – that is, the extent to which action potentials in the pulvinar were synchronized with the V4 LFP in various frequency bands – as a function of time in the trial. **D** and **F** are aligned relative to the cue (black vertical line) and **E** and **G** relative to the array onset; **D** and **E** when attention was allocated inside the receptive field and **F** and **G** when it was allocated elsewhere. **H** and **I** show spike–field coherence averaged across the delay period between pulvinar and V4 and between pulvinar and TEO. Source: Saalmann, Pinsk, Wang, Li, and Kastner, 2012. Reproduced with permission of AAAS.

dorsolateral PFC whether to maintain or switch behavioral strategies. The dmFC monitors performance, processes information about conflict and errors, and likely acts as a governor on PFC activity. The implementation of cognitive control involves the synchronization between regions that are, or will

be, processing prioritized information. The pulvinar nucleus of the thalamus effects this synchronization in the ventral visual stream, and for other systems the same may be true of how other higher-order thalamic nuclei influence oscillatory synchrony within the cortical territories to which they connect.

END-OF-CHAPTER QUESTIONS

1. Name two or three examples from your daily life that illustrate failures of cognitive control. In each case what was it about the behavior that you *did* produce that resulted in its being expressed rather

than the behavior that you should have produced? (Need a hint to get you started? Have you ever done anything “klutzy” while walking and texting on your mobile phone?)

2. What brain region is most strongly associated with cognitive control? Might there be sub-regional organization such that different aspects of control are preferentially carried out by different subregions?
3. Define a perseverative error on the WCST. What factors explain this distinctive failure of cognitive control? What are two possible mechanisms whereby perseverative errors are prevented in the normally functioning nervous system?
4. The study of PFC-lesioned patients with ERP has revealed two distinct loci of impaired neuronal functioning during an STM task – what are they?
5. How can models of PFC-based cognitive control rule out the logical problem of a PFC-based homunculus who decides when it's time to apply cognitive control?
6. Define the reward prediction error signal: what circumstances give rise to it and what brain system generates it?
7. Summarize three explanations for performance monitoring-related activity in dorsomedial PFC.
8. Why is interareal oscillatory synchrony important for the implementation of cognitive control? Summarize one mechanism for influencing inter-areal oscillatory synchrony.

REFERENCES

- Botvinick, Matthew M., Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen. 2001. "Conflict Monitoring and Cognitive Control." *Psychological Review* 108 (3): 624–652. doi: 10.1037/0033-295X.108.3.624.
- Chao, Linda L. and Robert T. Knight 1998. "Contribution of Human Prefrontal Cortex to Delay Performance." *Journal of Cognitive Neuroscience* 10 (2): 167–177. doi: 10.1162-089892998562636.
- Downey, Philip. 2006. "Profile of Ranulfo Romo." *Proceedings of the National Academy of Sciences of the United States of America* 103 (39): 14263–14265. doi: 10.1073/pnas.0607434103.
- Falkenstein, Michael, Joachim Hohnsbein, Jörg Hoormann, and L. Blanke. 1991. "Effects of Cross-Modal Divided Attention on Late ERP Components: II. Error Processing in Choice Reaction Tasks." *Electroencephalography and Clinical Neurophysiology* 78 (6): 447–455.
- Gehring, William J., Brian Goss, Michael G. H. Coles, David E. Meyer, and Emanuel Donchin. 1993. "A Neural System for Error Detection and Compensation." *Psychological Science* 4 (6): 385–390. doi: 10.1111/j.1467-9280.1993.tb00586.x.
- Gehring, William J., Yanni Liu, Joseph M. Orr, and Joshua Carp. 2012. "The Error-Related Negativity (ERN/Ne)." In *The Oxford Handbook of Event-Related Potential Components*, edited by Steven J. Luck and Emily S. Kappenman, 231–291. New York: Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0120.
- Hamker, Fred H. 2005. "The Reentry Hypothesis: The Putative Interaction of the Frontal Eye Field, Ventrolateral Prefrontal Cortex, and Areas V4, IT for Attention and Eye Movement." *Cerebral Cortex* 15 (4): 431–447. doi: 10.1093/cercor/bhh146.
- Holroyd, Clay B., and Michael G. H. Coles. 2002. "The Neural Basis of Human Error Processing: Reinforcement Learning, Dopamine, and the Error-Related Negativity." *Psychological Review* 109 (4): 679–709. doi: 10.1037/0033-295X.109.4.679.
- Knight, Robert T., and Mark D'Esposito. 2003. "Lateral Prefrontal Syndrome: A Disorder of Executive Control." In *Foundations of Cognitive Neuroscience* edited by Mark D'Esposito, 259–280. Cambridge, MA: MIT Press.
- Lhermitte, François, Bernard Pillon and M. Serdaru. 1986. "Human Autonomy and the Frontal Lobes. Part I: Imitation and Utilization Behavior: A Neuropsychological Study of 75 Patients." *Annals of Neurology* 19 (4): 326–334. doi: 10.1002/ana.410190404.
- Malmö, Robert B. 1942. "Interference Factors in Delayed Response in Monkeys after Removal of Frontal Lobes." *Journal of Neurophysiology* 5 (4): 295–308.
- Miller, Earl K., and Timothy J. Buschman. 2013. "Cortical Circuits for the Control of Attention." *Current Opinion in Neurobiology* 23 (2): 216–222. doi: 10.1016/j.conb.2012.11.011.
- Miller, Earl K., and Jonathan D. Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience* 24: 167–202.
- Milner, Brenda. 1963. "Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes." *Archives of Neurology* 9 (1): 90–100. doi: 10.1001/archneur.1963.00460070100010.
- Montague, P. Read, Peter Dayan, and Terrence J. Sejnowski. 1996. "A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning." *Journal of Neuroscience* 16 (5): 1936–1947.

- O'Reilly, Randall C., and Yuko Munakata. 2000. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA, and London: MIT Press.
- Picard, Nathalie, and Peter L. Strick. 1996. "Motor Areas of The Medial Wall: A Review of their Location and Functional Activation." *Cerebral Cortex* 6: 342–353. doi: 10.1093/cercor/6.3.342.
- Ridderinkhof, K. Richard, Markus Ullsperger, Eveline A. Crone, and Sander Nieuwenhuis. 2004. "The Role of the Medial Frontal Cortex in Cognitive Control." *Science* 306 (5695): 443–447. doi: 10.1126/science.1100301.
- Rougier, Nicolas P., David C. Noelle, Todd S. Braver, Jonathan D. Cohen, and Randall C. O'Reilly. 2005. "Prefrontal Cortex and Flexible Cognitive Control: Rules without Symbols." *Proceedings of the National Academy of Sciences of the United States of America* 102 (20): 7338–7343. doi: 10.1073/pnas.0502455102.
- Saalmann, Yuri B., Mark A. Pinsk, Liang Wang, Xin Li, and Sabine Kastner. 2012. "The Pulvinar Regulates Information Transmission Between Cortical Areas Based on Attention Demands." *Science* 338 (6095): 753–756. doi: 10.1126/science.1223082.
- Schultz, Wolfram. 2006. "Behavioral Theories and the Neurophysiology of Reward." *Annual Review of Psychology* 57: 87–115. doi: 10.1146/annurev.psych.56.091103.070229.
- Stedron, Jennifer Merva, Sarah Devi Sahni, and Yuko Munakata. 2005. "Common Mechanisms for Working Memory and Attention: The Case of Perseveration with Visible Solutions." *Journal of Cognitive Neuroscience* 17 (4): 623–631. doi: 10.1162/0898929053467622.
- Weigl, Egon. 1941. "On the Psychology of So-Called Processes of Abstraction." Translated by Margaret J. Rioch. *Journal of Abnormal and Social Psychology* 36 (1): 3–33. doi: 10.1037/h0055544. Originally published 1927 as "Zur psychologie sogenannter Abstraktionsprozesse." *Zeitschrift Fur Psychologie* 103: 2–45.

OTHER SOURCES USED

- Diamond, Adele, and Patricia S. Goldman-Rakic. 1989. "Comparison of Human Infants and Rhesus Monkeys on Piaget's A-Not-B Task: Evidence for Dependence on Dorsolateral Prefrontal Cortex." *Experimental Brain Research* 74 (1): 24–40. <http://www.devocogneuro.com/Publications/Diamond&Goldman-RakicAB.pdf>.
- Durstewitz, Daniel, Jeremy K. Seamans, and Terrence J. Sejnowski. 2000. "Dopamine-Mediated Stabilization of Delay-Period Activity in a Network Model of Prefrontal Cortex." *Journal of Neurophysiology* 83 (3): 1733–1750.
- Eling, Paul, Kristianne Derckx, and Roald Maes. 2008. "On the Historical and Conceptual Background of the Wisconsin Card Sorting Test." *Brain and Cognition* 67 (3): 247–253. doi: 10.1016/j.bandc.2008.01.006.
- Falkenstein, Michael, Joachim Hohnsbein, and Jörg Hoormann. 1990. "Effects of Errors in Choice Reaction Tasks on the ERP under Focused and Divided Attention." In *Physiological Brain Research*, vol. 1, edited by Cornelis Henri Marie Brunia, Anthony W. K. Gaillard, and A. Kok, 192–195. Tilburg, NL: Tilburg University Press.
- Grant, David A., and Esta Berg. 1948. "A Behavioral Analysis of Degree of Reinforcement and Ease of Shifting to New Responses in a Weigl-Type Card-Sorting Problem." *Journal of Experimental Psychology* 38 (4): 404–411. doi: 10.1037/h0059831.
- Harlow, Harry F., and J. Dagnon. 1943. "Problem Solution by Monkeys following Bilateral Removal of the Prefrontal Areas. I. The Discrimination and Discrimination-Reversal Problems." *Journal of Experimental Psychology* 32 (4): 351–356. doi: 10.1037/h0058649.
- Lhermitte, François. 1986. "Human Autonomy and the Frontal Lobes. Part II: Patient Behavior in Complex and Social Situations: The 'Environmental Dependency Syndrome.'" *Annals of Neurology* 19 (4): 335–343. doi: 10.1002/ana.410190405.
- Lisman, John E., Jean-Marc Fellous, and Xiao-Jing Wang. 1998. "A Role for NMDA-Receptor Channels in Working Memory." *Nature Neuroscience* 1 (4): 273–275. doi: 10.1038/1086.
- MacLeod, Colin M. 1991. "Half a Century of Research on the Stroop Effect: An Integrative Review." *Psychological Bulletin* 109 (2): 163–203. doi: 10.1037/0033-2909.109.2.163.
- Romo, Ranulfo, and Wolfram Schultz. 1990. "Dopamine Neurons of the Monkey Midbrain: Contingencies of Responses to Active Touch During Self-Initiated Arm Movements." *Journal of Neurophysiology* 63 (3): 592–606.
- Schultz, Wolfram, and Ranulfo Romo. 1990. "Dopamine Neurons of the Monkey Midbrain: Contingencies of Responses to Stimuli Eliciting Immediate Behavioral Responses." *Journal of Neurophysiology* 63 (3): 607–624.
- Settlage, Paul, Myra Zable, and Harry F. Harlow. 1948. "Problem Solution by Monkeys following Bilateral

- Removal of the Prefrontal Areas: VI. Performance on Tests Requiring Contradictory Reactions to Similar and to Identical Stimuli." *Journal of Experimental Psychology* 38 (1): 50–65. doi: 10.1037/h0054430.
- Skinner, James E., and Charles D. Yingling. 1976. "Regulation of Slow Potential Shifts in Nucleus Reticularis Thalami by the Mesencephalic Reticular Formation and the Frontal Granular Cortex." *Electroencephalography and Clinical Neurophysiology* 40 (3): 288–296. doi: 10.1016/0013-4694(76)90152-8.
- Yingling, Charles D., and James E. Skinner. 1976. "Selective Regulation of Thalamic Sensory Relay Nuclei by Nucleus Reticularis Thalami." *Electroencephalography and Clinical Neurophysiology* 41 (5): 476–482. doi: 10.1016/0013-4694(76)90059-6.
- Zable, Myra, and Harry F. Harlow. 1946. "The Performance of Rhesus Monkeys on Series of Object-Quality and Positional Discriminations and Discrimination Reversals." *Journal of Comparative Psychology* 39 (1): 13–23. doi: 10.1037/h0056082.

FURTHER READING

- Knight, Robert T., and Mark D'Esposito. 2003. "Lateral Prefrontal Syndrome: A Disorder of Executive Control." In *Neurological Foundations of Cognitive Neuroscience*, edited by Mark D'Esposito, 259–280. Cambridge, MA: MIT Press.
An authoritative review, from two behavioral neurologists, of the clinical presentation, and theoretical interpretations, of patients with damage to the lateral PFC.
- Miller, Earl K., and Jonathan D. Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience* 24: 167–202.
Paper that really set the agenda for cognitive control research for the first decade of the new millennium.
- Miller, Earl K., and Timothy J. Buschman. 2013. "Cortical Circuits for the Control of Attention." *Current Opinion in Neurobiology* 23 (2): 216–222. doi: 10.1016/j.conb.2012.11.011.
Summary of evidence, from human and monkey research, of the key role for interareal oscillatory synchrony as a mechanism for cognitive control.
- Passingham, Richard E., and Steven P. Wise. 2012. *The Neurobiology of the Prefrontal Cortex*. Oxford: Oxford University Press.
A systematic, focused, theory of frontal-lobe function, including a unique and provocative evolutionary focus, from two leading students of its anatomy, physiology, and function.
- Stuss, Donald T., and Robert T. Knight, eds. 2013. *Principles of Frontal Lobe Function*, 2nd ed. New York: Oxford University Press.
An exhaustive collection of chapters from many of the world's authorities on the myriad functions of the PFC.
- Wessel, Jan R. 2012. "Error Awareness and the Error-Related Negativity: Evaluating the First Decade of Evidence." *Frontiers in Human Neuroscience* 6: 88. doi: 10.3389/fnhum.2012.00088.
Review of the nuanced and fascinating literature on the role of awareness in performance monitoring and the ERN.

CHAPTER 15

DECISION MAKING

KEY THEMES

- Three classes of decision that are each studied differently, and may be differently implemented by the brain, are perceptual decisions, value-based decisions (the focus of neuroeconomics), and foraging decisions.
- Microstimulation of MT influences the decisions that monkeys make about the visually perceived direction of motion.
- Signal detection theory and drift-diffusion modeling interprets the activity of LIP neurons as representing the build-up of evidence in a decision variable.
- If the response is to be made manually, the build-up in the decision variable is seen in skeletomotor circuits.
- In a value-based oculomotor decision task, LIP neurons can represent the value of a saccade, raising the possibility that the representation of a motor value map, rather than a salience map, may be the fundamental function of the oculomotor system.
- Extracellular electrophysiology in the monkey and fMRI in the human implicate orbital and ventromedial PFC in the representation of the “common currency” that underlies behavioral economic theory.
- MVPA of fMRI scans of a free-choice button-pressing task suggests that seeds of the decision may be present in the brain as much as 10 seconds prior to a subject’s awareness of her conscious decision that she will push the button.

CONTENTS

KEY THEMES

BETWEEN PERCEPTION AND ACTION

PERCEPTUAL DECISION MAKING

Judging the direction of motion

Microstimulation of MT

LIP

Perceptual decision making in humans

VALUE-BASED DECISION MAKING

The influence of expected value on activity in LIP

Decision variables emergent from motor plans

Common currency in the omPFC

Transitivity, value transitivity, and menu

invariance

Evidence from the nonhuman primate

Evidence from humans

vmPFC or OFC?

Has neuroeconomics taught us anything about
the economics of decision making?

FORAGING

I CAN'T DECIDE!

NEXT STOP

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

BETWEEN PERCEPTION AND ACTION

The simplest sensorimotor event is the reflex – the patellar tendon of your knee is tapped by the doctor’s mallet and your leg kicks out. This response to a stimulus involves just one synapse in the CNS: the stretch receptor in your quadriceps muscle makes a synaptic connection, in the spinal cord, directly onto a motor neuron which, in turn, commands a contraction of the quadriceps muscle. This circuit is called a reflex arc, and its action does not entail any kind of decision. Once initiated, its outcome is known with certainty. It is a deterministic system.

For behaviors that qualify as cognitive, on the other hand, all actions can be said to be the result of a decision. Most are not deterministic: given the same stimulus inputs, an individual may respond to it in different ways, depending upon the context, on the current state of any of many systems in his brain, on recent history with the same or similar types of stimuli. This raises a question that will be implied, if not explicit, in all that we will consider in this chapter: Whereas the simple reflex can clearly be represented as *[stimulus → response]*, are cognitive behaviors better construed as *[stimulus → decision → response]*, or are they the result of more elaborate *[stimulus → response]* trajectories? That is, although decisions mediated by the brain involve many, many more synaptic links than do spinally mediated reflexes, and processes that are running in parallel, do the two differ only quantitatively, or are they fundamentally, qualitatively different?

There are three domains of decision making that we will consider here. **Perceptual decisions** are closely related to the processes of recognition – “snap” decisions by which we classify what kind of thing we are seeing (or hearing, or ...) and then take the appropriate action – *If that pedestrian is walking out into the street, I will apply my brakes; if she is standing still, I will continue driving past the crosswalk.* **Value-based decisions** are the purview of **neuroeconomics** – *Out of the (seemingly) hundreds of choices on the shelves, which bottle of red wine will I select to purchase and bring home to enjoy with dinner?* The third domain, **foraging decisions**, evoke (for me, at least) that timeless question

posed by the British punk rock band The Clash: *Should I stay or should I go? That is, Do I remain where I am to exploit the resources currently before me, or do I set out to explore new opportunities?*

Oh, right. And some readers may be wondering whether we are going to address the question of whether or not humans truly exercise free will. Well, were we to do so in this book, this would be the chapter in which to do it ...

We begin with perceptual decision making – was the overall direction of motion within that cloud of dots moving up or down? We’ll begin, as we have in other domains of cognitive neuroscience, by considering the “laws of ... optics ... and receptor binding.” From this starting point, how far will we get in our understanding of this behavior, and its neural underpinnings, before, as Wolfram Schultz admonished in the previous chapter, we need to turn to the behavioral sciences, particularly psychology and economics, for models that will help us understand even this simplest of decision-making behaviors?

PERCEPTUAL DECISION MAKING

Judging the direction of motion

Microstimulation of MT

Much later in this chapter, we will consider the neural bases of volitional action that reflects, as closely as possible, wholly internally generated decisions to act. Much more scientifically tractable, however, are situations in which decisions are made about sensory stimuli. This is because one can systematically vary properties of the inputs, and determine how this variation influences the decision, thereby gaining important knowledge about how the decision-making process works. A highly influential example of such an experiment, performed in the laboratory of neurophysiologist William Newsome, is illustrated in *Figure 15.1*. Newsome, Britten, and Movshon (1989) had previously studied the relationship between the activity of neurons in MT and a monkey’s decision about whether the overall, integrated motion signal from a cloud of randomly moving dots (referred to as the “global” motion of the cloud, to distinguish it from the “local” motion of any one dot) was in one

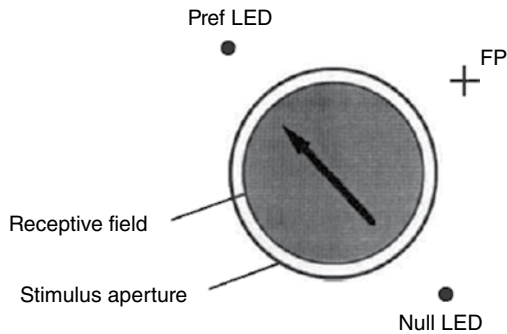


FIGURE 15.1 Influencing perceptual decisions with microstimulation. Diagram of the task, in which the directional tuning of an MT neuron was first determined (in this illustration, up and to the left), then the monkey performed several trials in which random dot motion (RDM) was presented in a stimulus aperture that covered the neuron's receptive field. For each, the animal made a forced choice as to whether the global direction of motion was toward the location marked "Pref LED," or toward the "Null LED," and indicated its decision with a saccade made to that LED. The strength of the motion signal was controlled by varying the percentage of dots all moving in the same direction, such that a 25% "correlation" meant that 25% of the dots were moving in the neuron's preferred ("Pref") direction, a -8% correlation that 8% of the dots were moving in the neuron's anti-preferred ("Null") direction. On a randomly determined subset of trials, microstimulation was delivered to the neuron. Source: Salzman, Britten, and Newsome, 1990. Reproduced with permission of Nature Publishing Group.

direction or another. Intriguingly, they found that the stimulus-related activity of some individual MT neurons was as good as, and in some cases even better than, the performance of the monkey! As a next step, to establish the causal role of MT neurons tuned to a specific direction of motion in the animal's decisions about motion in that direction, they performed the experiment illustrated in *Figure 15.1*.

The results were striking. Let's first consider just one data point in panel A of *Figure 15.2* – the *no stimulation* trials when the random dot motion (RDM) stimulus contained 0% correlated motion. On these trials, there was no global motion, and the animal decided *Pref* on roughly half the trials (to my eye it

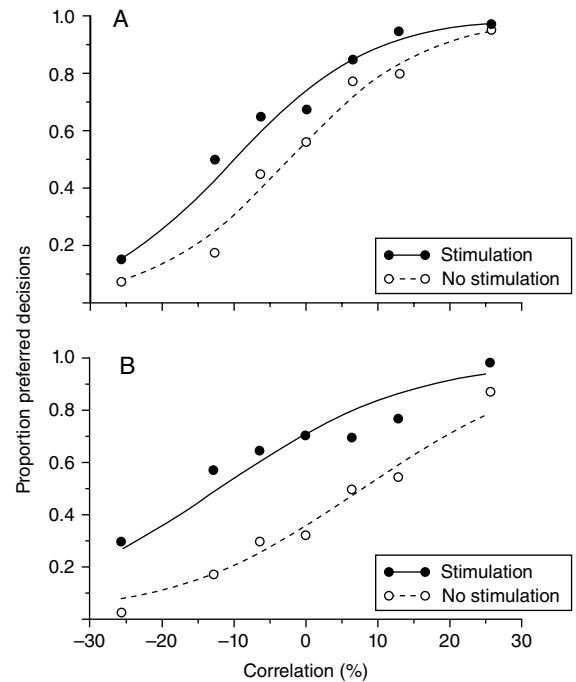


FIGURE 15.2 The results of varying the strength of the motion signal, and of microstimulation, at two stimulation sites within MT. The data are plotted as the animal's choice (expressed as the proportion of saccades made to the Pref LED) as a function of motion signal strength (Correlation). Source: Salzman, Britten, and Newsome, 1990. Reproduced with permission of Nature Publishing Group.

looks like 55% of these trials) and *Null* on the other half. Now look at its performance on the perceptually identical trials when microstimulation was delivered to an MT neuron tuned to detect motion in the *Pref* direction: on *stimulation* trials, the monkey decided *Pref* much more often, somewhere between 65% and 70% of the time. The same is true, qualitatively, at every level of stimulus signal strength, when the true global motion was in either direction.

When thinking about the effect of microstimulation for one stimulus type (e.g., 0% correlation), it's easier for me to think about it in terms of "How does microstimulation influence the monkey's decision?" Answer: it biases it to give more *Pref* responses. But this leads to additional questions whose answers are

Tip 15.1

Of course, we can't know what the monkey is perceiving when microstimulation biases its perceptual decisions, we can only infer this from the behavior. Framing the microstimulation effects in terms of what the animal perceives does seem like a reasonable supposition, though, when one considers, for example, that we do know that "macrostimulation," with transcranial magnetic stimulation, systematically causes human subjects to perceive phosphenes (*Research Spotlight 3.2*).

less straightforward. For example, does this mean that microstimulation makes the monkey perceive more of the stimuli as having motion in the preferred direction? Of course, we can't know this with any certainty (*Tip 15.1*), and so this makes characterization of the effects of microstimulation in terms of the animal's performance seem somehow subjective and less-than-satisfying. For this reason, even though on a stimulus-by-stimulus basis the effect of microstimulation is to move each data point "up" relative to the ordinate, a more useful way to quantify its effects is in terms of how it shifts the animal's psychometric curve along the abscissa. The reason is that this allows the effects of microstimulation to be quantified in terms of the physical properties of the stimulus that was varied in order to derive this curve. Thus, for the stimulation site depicted in the upper panel of *Figure 15.2*, microstimulation was equivalent to adding 7.7% dots moving in the preferred direction to each stimulus display. For the site in the lower panel, the effect was much larger, corresponding to a shift of 20.1% correlated dots in the preferred direction. (Note that, for both sites illustrated, particularly the one in the lower panel, there also seems to be some distortion to the shape of the psychometric function. What might explain this?)

Part of the importance of the Salzman, Britten, and Newsome (1990) study was that it prescribed a well-understood circuit in which perceptual decision making could be studied in greater depth: first, the microstimulation results indicated that MT is, indeed, causally involved in perceptual decision making (about motion); second, the task was, in effect, a visually

guided eye-movement task. "This link," as Gold and Shadlen (2007, p. 545) point out, "enables investigators to treat the decision as a problem of movement selection." Because the eye movement command would have to pass through FEF and/or SC, the logical next area to investigate in the perceptual decision-making process, one that links MT with these two oculomotor control centers, was LIP.

LIP

Because LIP is two synapses downstream from MT, the question is whether its contribution to the perceptual decision task reflects perceptual processing, decision making, or motor planning and execution. In keeping with the idea of treating the decision as a problem of movement selection, the procedure for studying neurons in LIP is to place one of the two saccade targets inside the neuron's motor field, (i.e., not the aperture displaying the RDM). With this procedure, it turns out to be fairly straightforward to rule out a perceptual role for LIP, simply by examining trials that present RDM containing 0% correlated motion. On roughly half of these trials, a saccade will be made into the motor field of the neuron being recorded from, and on the other half it will be made away from it. If the neuron's function were perceptual in nature, its response would be the same for both decisions, because on all trials the stimulus is identical. However, this is not what is observed. Instead, for the former, the neuron shows a gradual increase in firing rate, until the time of the saccade, and for the latter, it shows a suppressed level of activity until after the saccade. Interpretation of the ramp-up observed on trials when the saccade will come into the neuron's motor field, however, is more complicated. One possibility is that it corresponds to an accumulation of information related to the decision. The alternative is that it represents motor preparation, the decision having already been made elsewhere. Is it possible to rule out one of these two accounts?

Initial evidence that activity in LIP may correspond to the build-up of decision-related information comes from the rate of increase in activity that is observed as a function of stimulus strength: on trials in which the RDM contains a larger percentage of correlated motion, the increase in firing rate is steeper than it is on trials presenting weaker directional

information. Does this steeper increase in firing rate result in the animal reaching its decision more quickly? It turns out that it does, as can be determined by making a change to the experimental procedure: rather than having the animal make a response after a fixed-length, longish delay, allow it to respond as quickly as possible, incentivizing speed by providing larger rewards for quicker responses. The first thing one notices on the speeded variant of the task is that response time (RT) is slower for stimuli that provide weaker directional information. This pattern in the behavior is also seen in the activity of LIP neurons (Figure 15.3). From the perspective of signal-detection theory (SDT), this is the behavior that one would expect of a hypothesized **decision variable** that accumulates the information used to make a decision.

(Recall that we introduced SDT in the context of dual-trace models of LTM, *Chapter 11*. In particular, the build-up of evidence from a familiarity-based signal is reminiscent of our present topic.)

Regardless of stimulus strength, the theory predicts that the decision variable must achieve a certain value before a decision is made, and this would seem to correspond to when LIP neurons representing the decision (i.e., “Select T_{in} ” from Figure 15.3) reach a firing rate of 70 Hz. Because LIP neurons achieve this level of activity with RDM displays containing no global direction information, they (and the hypothesized decision variable) must be receiving additional information apart from what is fed forward from MT. SDT holds that this additional information corresponds to expectations that have been built up over

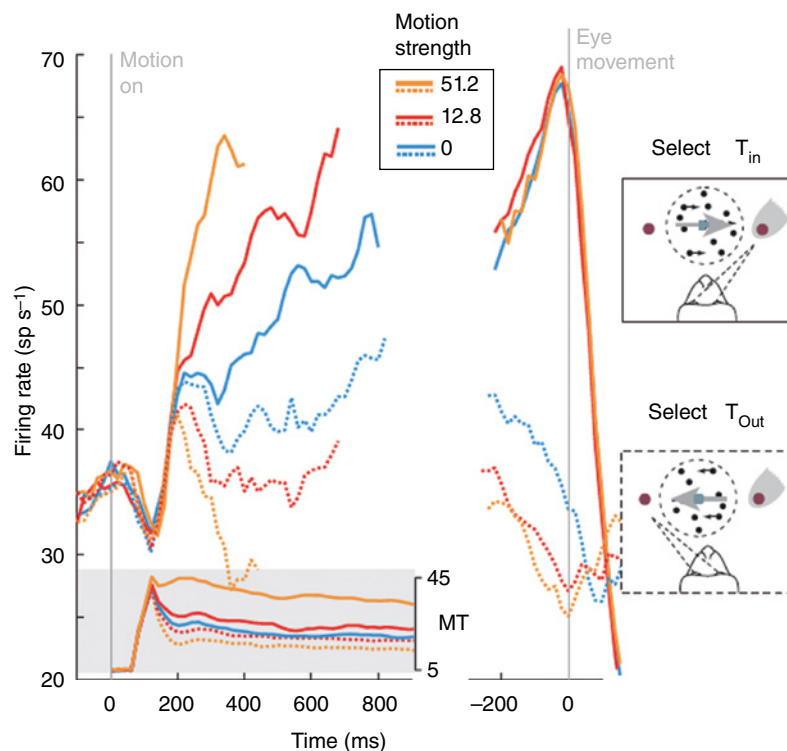


FIGURE 15.3 Activity in LIP, displayed as peristimulus time histograms (PSTH), in speeded RDM direction decision task, as a function of motion strength and whether the direction selected corresponds to a saccade made into the neuron’s motor field (“Select T_{in} ”; solid lines) or away from the neuron’s motor field (“Select T_{out} ”; dotted lines). Data on the left are aligned to trial onset, and on right are aligned to saccade onset. Inset, in shaded box in lower left, is activity in an MT neuron tuned for motion in the Select T_{in} direction. Source: Gold and Shadlen, 2007. Reproduced with permission of Annual Reviews.

past experience (“prior probabilities” in terms of Bayesian statistics), and information about the reward value of the choice. More generally, one might think of it in terms of the predictive coding framework introduced at the end of *Chapter 3* – where there is insufficient sensory evidence, behavior will be governed by the brain’s current model of the environment.

Methodology Box 15.1 will look more thoroughly (if not more formally) at the SDT approach to perceptual decision making, and *Research Spotlight 15.1* will detail an application of the theory to parse the contributions of MT and LIP to perceptual decision making. First, though, we need a quick refresher on the functional wiring diagrams of the cortico-basal

METHODOLOGY BOX

15.1 A signal detection theory-based framework for perceptual decision making

In the framework of Gold and Shadlen (2007), decision making is formalized in terms of SDT and the related theory of sequential analysis; we’ll abbreviate these as SDT. This methodology box will not present the formalisms (i.e., the math) underlying these theories, but will seek to convey the gist of how they are used to understand perceptual decision making.

“A decision,” write Gold and Shadlen (2007), “is a deliberative process that results in the commitment to a categorical proposition. An apt analogy is a judge or jury that must take time to weigh evidence for alternative interpretations and/or possible ramifications before settling on a verdict” (p. 536). The key factors to understand, from this perspective, are how the deliberative process works, and how it is translated into the commitment. In *Figure 15.3*, the neural correlate of the deliberative process is captured by the evolution of any two same-colored PSTHs, which are believed to reflect the temporal evolution of evidence for a *Select T_{in}* vs. a *Select T_{out}* decision. The sum of these two processes, running in parallel, is the decision variable, which Gold and Shadlen (2007) note, “is not tied to the (possibly fleeting) appearance of stimuli but spans the time from the first pieces of relevant information to the final choice” (p. 538). In this regard, the concept of the decision variable can be understood in relation to STM, as can the sustained, elevated activity in LIP illustrated in *Figure 15.3*.

The processes that contribute to deliberation, and their culmination in a commitment, can be quantified and modeled in drift-diffusion models (also called “diffusion-to-bound” models). Developed by experimental psychologists to account for behavior in speeded decision tasks, these are “mathematical models in which sensory information is integrated stochastically over time, resulting in a **random walk** of an abstract variable toward a preset decision threshold. A decision is made when the random walker reaches the threshold” (Lo and Wang, 2006, p. 956). *Figure 15.4* illustrates how computational neuroscientist Xiao-Jing Wang and his colleagues have incorporated these formalisms into a neurobiologically inspired model of decision making. The model implements evidence accumulation via “slow reverberatory network dynamics mediated by NMDA receptors” (p. 956; *Tip 15.2*). Ultimately, the decision is executed by burst neurons in the superior colliculus (SC). The commitment is achieved through a dynamic interplay between direct cortical input onto the SC and a basal ganglia-mediated input from cortex to caudate nucleus to SNpr to SC. The location of the threshold (i.e., how high the firing rate in LIP needs to get) is determined by the strength of synapses from cortex onto the caudate nucleus. This is because although LIP has a direct input to the SC, the excitatory influence of this input is counteracted by the tonic inhibition from the SNpc.

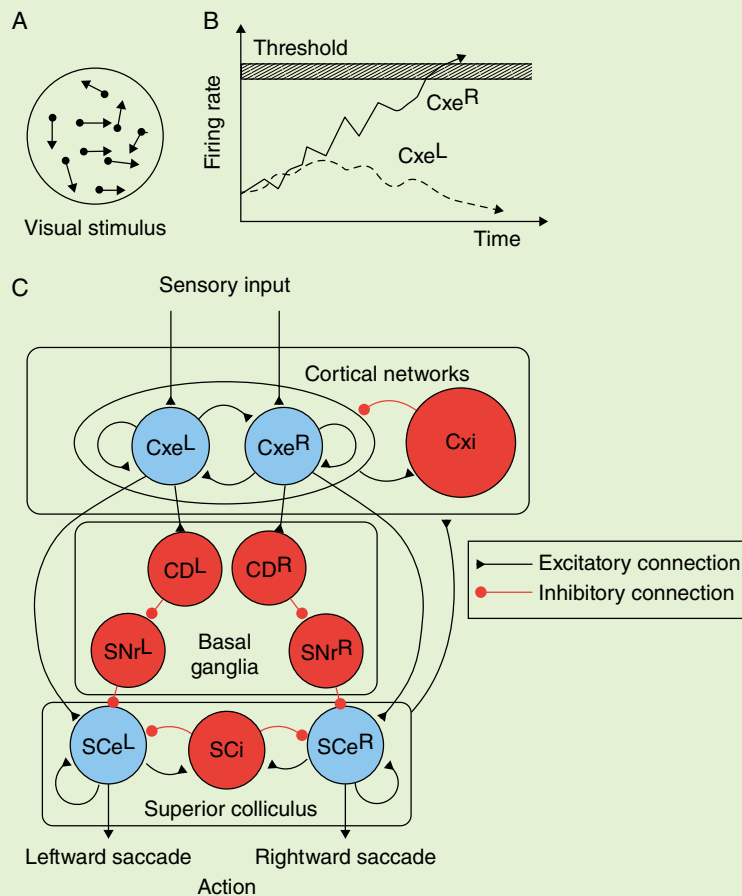
Methodology Box 15.1 *Continued*

FIGURE 15.4 “Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks.”

A. Schematic of the RDM display. **B.** A schematic representation of the random walk followed by two cortical populations (Cxe^L and Cxe^R), which we can think of as LIP neurons representing $Select\ T_{in}$ and $Select\ T_{out}$ from Figure 15.3. Commitment occurs (i.e., a saccade is generated) as soon one of the population firing rates reaches the threshold. **C.** Schematic model architecture. Neural pools in the cortical network integrate sensory information and also compete against each other. They project to both the superior colliculus (SCe^L and SCe^R) and the caudate nucleus (CD^L and CD^R). Blue, populations of excitatory neurons; red, populations of inhibitory neurons. Each population is simulated by noisy spiking leaky integrate-and-fire neurons. Source: Lo and Wang, 2006. Reproduced with permission of Nature Publishing Group.

In summary, there are two principal parameters in the drift-diffusion model that influence the speed, and the accuracy, of the decision-making process: the “drift rate,” which corresponds to the slope of the random walker; and the decision boundary, which sets the threshold that, once achieved, triggers the commitment to a decision.

Tip 15.2

Perhaps NMDA channel-mediated sustained activity in a recurrent network isn't a unique property of the PFC, as was suggested in *Chapter 14*?

ganglia-thalamic system, particularly of the oculomotor loop, and its relation to the superior colliculus. Cortex projects to the caudate nucleus, which then splits this signal into a direct and an indirect pathway. Most relevant to our present focus is the direct pathway's inhibitory projection from the caudate nucleus to the substantia nigra pars reticulata (SNpr; *Figure 7.2* and *Figure 8.1*). The SNpr, like the internal segment of the globus pallidus (GPi) is tonically active, and inhibitory. Thus, when a channel within the oculomotor circuit of the basal ganglia is engaged – a channel corresponding to a specific portion of retinotopic space – the inhibitory clamp of the SNpr on the superior colliculus is temporarily lifted, and a saccade can be triggered. Now we're ready for *Methodology Box 15.1* and *Research Spotlight 15.1*.

It's worthy of note that the mechanisms underlying visuoperceptual decision making seem to be the very same – both at the level of the circuitry and at the level of the function – as those involved in eye movement control. In effect, to understand this aspect of “high-level cognition,” what we have done is incorporate concepts from SDT into what we already know about how oculomotor control works. Thus, at least for the type of behavior that we've studied thus far, it seems that there is not a distinct decision-making system. Rather, decision-making functions are emergent from the oculomotor system. To see if this generalizes, we'll next look at data from a task requiring a comparable perceptual decision, but requiring responses with a different motor effector system. And, for good measure, we'll make it one that tests a different species, and that uses a different measure of brain activity.

Perceptual decision making in humans

Perceptual decision making has also been studied extensively in humans, for decades with behavioral studies, and also with many neurophysiological measures. The study summarized here was selected because it entailed the by-now familiar decision about RDM,

but required manual button-press responses, rather than eye movements. It was performed by a group of German scientists based at institutions in Germany and in the Netherlands. The two senior authors, Andreas Engel and Pascal Fries, have both made pioneering contributions to our understanding of the functional and cognitive roles of oscillatory dynamics in the brain. Not surprisingly, then, oscillatory power is the principal dependent measure used.

The task required subjects to indicate, on each trial, whether or not there was coherent motion embedded in the RDM display. Magnetoencephalographic (MEG) measurements were acquired during task performance, and the analyses focused on spectral transforms of the data. Responses were lateralized – one hand for *yes*, the other for *no* (counterbalanced across subjects) – to allow for clear isolation of *yes* hand-related vs. *no* hand-related preparatory and response activity. The researchers focused on two broad frequency bands, beta (which they operationalized as 12–36 Hz) and gamma (64–100 Hz). Prior research had indicated that limb movements are accompanied by suppression activity in the beta band and enhancement of activity in the gamma band, both typically stronger in the motor cortex contralateral than ipsilateral to the side of the movement (see *Figure 7.6*). These dynamics allowed for the computation of choice probability – a statistical estimate of how well a hypothetical ideal observer could predict a subject's impending choice based on differences in signal between the two hemispheres. (Conceptually, not unlike how MVPA performance is evaluated.)

The findings, summarized in *Figure 15.6*, were consistent with the conclusion that we drew tentatively at the end of the previous section, which is that the evolution of the decision variable can be observed in the pre-movement activity of the system controlling the motor effector that will make the response. In both the beta band and the gamma band, and in both ROIs, choice probability estimates began to ramp up as early as 1.5 seconds prior to the button press. In a further analysis, the authors leveraged a previous finding of theirs, in which they had found that gamma-band power in MT scales monotonically with motion coherence. They reasoned that, if the motor cortex oscillatory dynamics that they had observed were truly reflective of the evolution of the decision variable,

RESEARCH SPOTLIGHT

15.1 Differential effects of microstimulation of MT and LIP during perceptual decision making (Hanks, Ditterich, and Shadlen, 2006)

There is an additional level of detail to SDT that was not included in *Methodology Box 15.1*: SDT makes a distinction between “momentary evidence” and “accumulated evidence.” In the study spotlighted here, Hanks, Ditterich, and Shadlen (2006) hypothesized that, in the RDM direction decision task, MT is a source of momentary evidence, whereas LIP represents accumulated evidence for a decision. To illustrate, we can return again to *Figure 15.3*. Momentary evidence would be what is being generated by neurons in MT for the duration of time that the RDM stimulus is on the screen. In paradigms with enforced delay periods separating the RDM stimulus and the response, MT activity returns to baseline after the offset of the RDM, meaning that the stimulus-driven momentary evidence drops off to 0 during the delay period. Accumulated evidence, however, represented in LIP firing rates, acts as an integrator, and so continues to build, even after stimulus offset.

To test this hypothesis, Hanks, Ditterich, and Shadlen (2006) took a page from the book of Shadlen’s postdoctoral mentor, William Newsome, and compared the effects of microstimulation delivered to MT vs. to LIP. Importantly, they fit the behavioral data from the monkeys to a drift-diffusion model, as summarized in *Methodology Box 15.1*. If MT does, indeed, represent momentary evidence, an increase in its drive was expected to steepen the drift rate, thereby biasing decisions in favor of the direction preferred by the stimulated neurons. And if LIP represents the decision variable, stimulating it was expected to produce an offset in the decision variable – a “DC shift-like” jump – in the direction of one of the two decision bounds. (Note that this is mathematically equivalent to moving one of the bounds closer to the 0 point from which the decision variable begins its random walk at the start of each trial.)

The principal empirical effects, illustrated in *Figure 15.5* panels B and C, were that LIP microstimulation had a proportionally greater effect on RT than it did on choice, whereas MT microstimulation had a strong biasing effect on choice (replicating Salzman, Britten, and Newsome, 1990), but only noisy effects on RT. Fitting the data to the drift-diffusion model indicated that MT microstimulation changed the drift rate to a degree that would have been produced by adding 9.3% correlated motion¹ to the RDM display. With regard to decision variable offset, however, the MT microstimulation had no significant effect. For LIP, the effect was opposite: the effect of LIP microstimulation on drift rate did not differ from 0, but its effect on the decision variable was to offset it 16.9% closer to the decision boundary. Similar studies suggest similar functions for the FEF.

From the perspective of the Lo and Wang (2006) model, one might speculate that microstimulation of LIP produced a transient strengthening of LIP’s synaptic connections with the caudate nucleus; a strengthening that is not produced by a stronger afferent drive from extrastriate visual areas.

¹Note that in *Figure 15.5* the term “coherence” is used. This is the more standard term (rather than “correlation”) for describing the percentage of dots moving in the same direction in a RDM display.

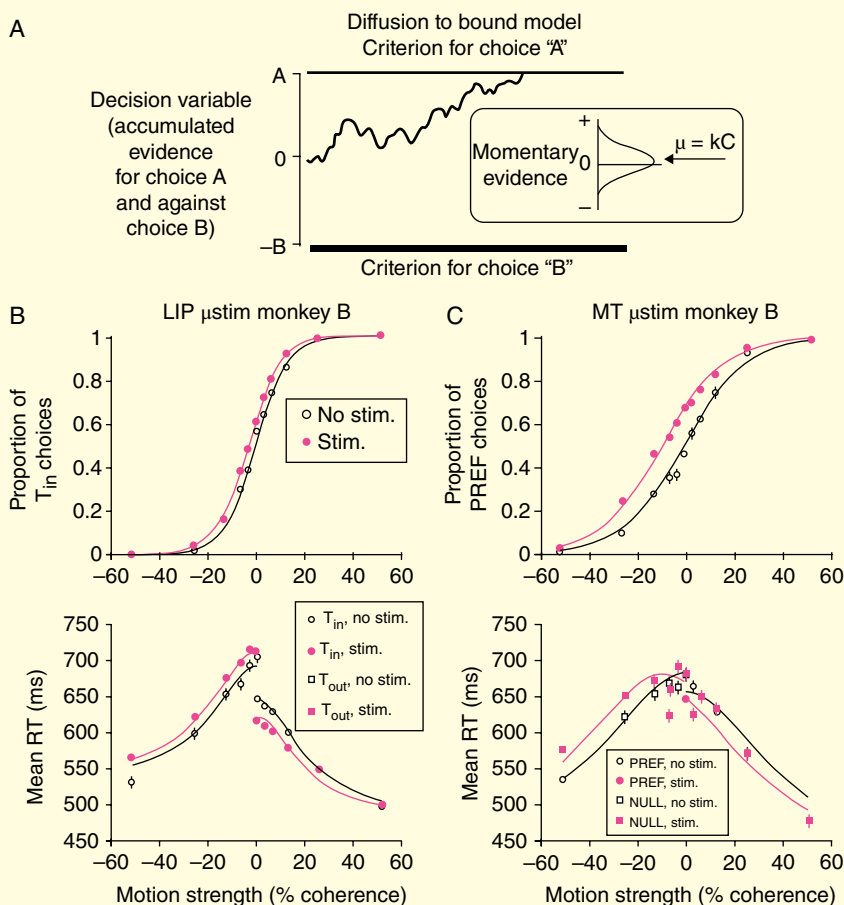
Research Spotlight 15.1 *Continued*

FIGURE 15.5 A drift-diffusion model, and model fits to data, of the effects of microstimulation in LIP and MT. **A** illustrates the behavior of the drift-diffusion (or “diffusion-to-bound”) model on a single hypothetical trial. It differs from the model in *Figure 15.4*, panel B, only in that *15.4*, panel B, portrays a race between two processes trying to achieve the same threshold, and this figure portrays the two as racing in opposite directions, each toward its threshold. The two are mathematically equivalent. The inset illustrates that momentary evidence is stochastic, thereby making the drift a random walk, and pulled on a time increment-by-time increment basis from a normal distribution centered on a mean whose value is represented by (μ), a value that is linearly related to motion strength (C ; k is a free parameter that scales the motion appropriately). **B**. Effects of microstimulation of LIP. Conventions for decision plot are the same as from *Figure 15.2*. Effect of microstimulation on choice is negligible, but on RT is to speed RTs for T_{in} trials and to slow them for T_{out} trials. (Overall difference between T_{in} and T_{out} is due to unexplained, but constant, bias in this animal during these sessions.) **C**. Effects of microstimulation of MT, showing large influence on choice, but no reliable effect on RT. Source: Hanks, Ditterich, and Shadlen, 2006. Reproduced with permission of Nature Publishing Group.

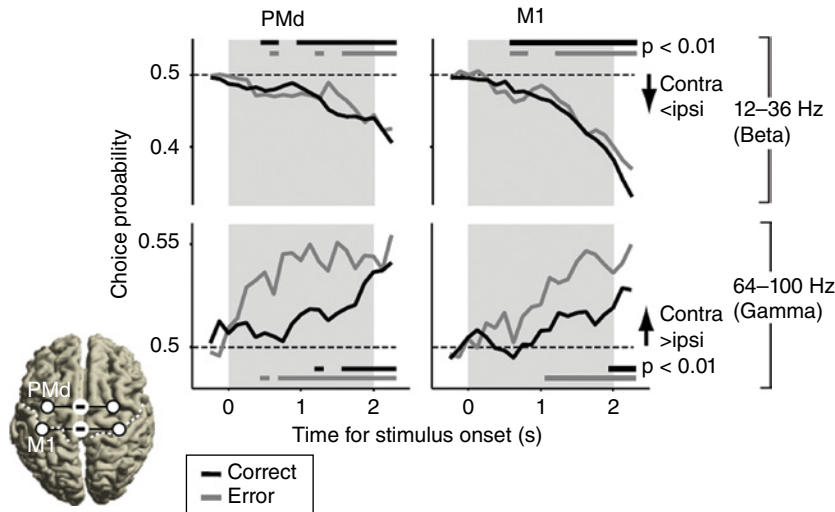


FIGURE 15.6 Build-up of choice-predictive activity in human motor cortex during perceptual decision making. Superimposed on the top-down view of a brain are the cortical regions of interest in the hand area of M1 and in premotor cortex (PMd). (The white dotted line traces the central sulcus.) The plots show choice probability (see text), estimated from two ROIs and from two frequency bands, ramping toward a decision in each cell. The shaded region indicates stimulus presentation; bars along the top and bottom indicate periods during which choice probability estimates are statistically significant. Source: Donner, Siegel, Fries, and Engel, 2009. Reproduced with permission of Elsevier.

trial-by-trial choice-predictive activity from these ROIs should relate to trial-to-trial fluctuations of the temporal integral of gamma-band activity in MT (i.e., to the strength of the momentary evidence). This prediction was borne out, thus reinforcing their interpretation of activity from precentral ROIs as reflecting the integration of sensory evidence provided by MT. (*Web Clip 15.1* illustrates how perceptual decision making can improve with intensive training.)

VALUE-BASED DECISION MAKING

Most decision theories – from expected utility theory in economics (von Neumann and Morgenstern, 1944) to prospect theory in psychology (Kahneman and Tversky, 1979) to reinforcement learning theories in computer science (Sutton and Barto, 1998) – share a core conclusion. Decision makers integrate the various dimensions of an option into a single measure of its idiosyncratic subjective value and then choose the option that is most valuable. Comparisons between different kinds of

options rely on this abstract measure of subjective value, a kind of “common currency” for choice. That humans can in fact compare apples to oranges when they buy fruit is evidence for this abstract common scale. (Kable and Glimcher, 2009, p. 734)

Thus, the quest for the common currency will be an important theme underlying this section of the chapter.

From the outset, value-based decision making would seem to be different in kind from the perceptual decision making that we just covered. Whereas the latter entails a decision about the objective state of the world – either the dots were drifting up or they were drifting down – the former is inherently subjective – if you randomly select two people and present them with the same choice, one can prefer the apple and one can prefer the orange, and neither of these choices can be said to be wrong. This is also called economic choice, as the quote from Kable and Glimcher indicates, and economists have studied this aspect of behavior for decades. In the context of this textbook, what we’ll be discussing has come to be referred to as neuroeconomics. To get started, we’ll

begin with an experimental procedure that shares many surface characteristics with the perceptual decision-making studies that we have just reviewed, and that measures neural activity in familiar territory.

The influence of expected value on activity in LIP

This study by Michael Platt and Paul Glimcher (1999), two prominent figures in neuroeconomics research, was designed from the perspective that “nearly all theories require the decision-maker to have some knowledge of two environmental variables: the gain expected to result from an action and the probability that the expected gain will be realized.” In expected value theory, for example, “a rational decision-maker should multiply expected gain by the probability of gain to establish an expected value for each course of action, and then should choose the option with the highest expected value” (p. 233). To evaluate this idea, they trained monkeys to fixate centrally while two target stimuli were presented in the periphery, one always in the movement field of the LIP neuron being recorded. Upon offset of the fixation point, the monkey was rewarded for making a saccade to either target (i.e., it was a “free choice”). For the first block of trials, the amount of juice reward associated with each of the targets (i.e., the gain) was identical. Subsequently,

the gain associated with each target was modified, such that selection of one yielded 0.15 ml of juice, and the other 0.05 ml. These gains were switched between the targets, on a block-by-block basis, such that across several blocks of trials each LIP neuron was recorded under perceptually identical conditions, but when the expected gain of selecting the target in its movement field was relatively high or relatively low.

The results, illustrated in *Figure 15.7.A*, revealed a marked difference in response profile of LIP neurons for high-gain vs. low-gain decisions, with the former showing elevated activity even prior to target onset. (One can speculate that, even prior to target onset, the animal had “already made up its mind.”) Note, however, that this was no longer true at the time of response. It is also important to note that the monkeys did not simply always choose the target with the highest expected gain. Rather, they followed the “matching law,” a tendency for animals to allocate their choices in direct proportion to the rewards provided by each choice. (That is, they chose 0.15 ml on 75% of the trials, and 0.05 ml. on 25% of the trials; *Tip 15.3.*) As a result, Platt and Glimcher were able to compute an estimate of the animal’s valuation of each response on a trial-by-trial basis (i.e., that response’s expected value) by applying a formula from a formal theory of behavioral decision making – melioration theory – that took into account the animal’s choices

FIGURE 15.7 Activity of a single LIP neuron (identified as CH980304) from kinematically identical saccades made into its movement field, but at two levels of expected gain (0.15 ml vs. 0.05 ml of juice). Source: Platt and Glimcher, 1999. Reproduced with permission of Nature Publishing Group.

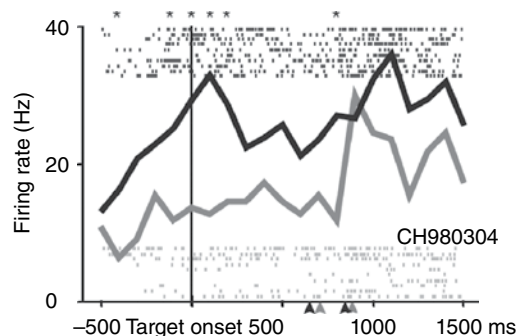


FIGURE 15.7.A Average activity, time-locked to target onset, for high (black) and low (gray) expected-value trials. Arrows along the timeline indicate, sequentially, and for each trial type, the average time of fixation offset and of saccade initiation. Asterisks indicate 100 ms bins in which firing rate differed significantly at the two levels of expected gain. Note that fixation onset preceded target onset by 500 ms.

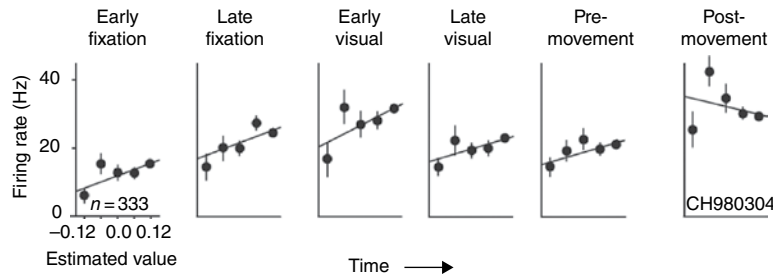


FIGURE 15.7.B Activity from the same neuron, divided into 200ms samples, and regressed against expected value, for movements into the neuron's response field.

Tip 15.3

Glimcher (2002) has noted that such matching behavior is optimal for an individual animal that is foraging with a group, and therefore competing with others for access to rewards of different values. Could it be that, even when alone in a dark, sound-attenuated recording chamber, the brains of Platt and Glimcher's (1999) monkeys were operating under an "evolutionary assumption that they [were] competing with other monkeys" (p. 328)?

over the 10 preceding trials. *Figure 15.7.B* shows that activity in this LIP neuron related to expected value during all epochs of the trial, including the two that preceded target onset, with the exception of the "post-movement" epoch (which was actually time-locked to the initiation of the saccade).

Decision variables emergent from motor plans

Thus, to recap what we've learned in LIP, Shadlen and colleagues have shown that neurons in this region participate in the accumulation of evidence – the deliberation – toward a perceptual decision (*Figure 15.3* and *Figure 15.5*). In the study just summarized, Platt and Glimcher (1999) have shown that, in a free-choice task that varies gains over time, LIP neurons can also represent the value of eye movements. One consequence of the latter, and related, findings is that some have proposed a reframed conception of the frontoparietal oculomotor complex. Rather than this

system representing a salience map (*Chapter 8*) – which is a fundamentally attention-based construal of brain function – perhaps it represents a "saccade value" map – a fundamentally neuroeconomic construal of brain function. In-depth consideration of the relative merits of these two perspectives is beyond the scope of this book. However, there is one additional conclusion from this literature that will help us wrap up where we've been and to push off toward our next destination. Studies of value-based decision making in parietal cortex have established that the representations of value found in LIP are limited to the value of decisions about where to move the eyes. Thus, by analogy to what we saw with perceptual decision making, value can be represented in the networks that will effect the value-based action. A further implication of this fact is that LIP does not seem to represent the common currency that is, in some sense, a Holy Grail of neuroeconomics. For that, neuroeconomic Sir Galahads have been more successful in the orbital and ventromedial PFC (sometimes summarized as "orbito-medial PFC" (omPFC)).

Common currency in the omPFC

It is well established in behavioral neurology that lesions of the omPFC impair choice behavior in many domains, including eating disorders, abnormal gambling, and, more generally, erratic decision making. In the monkey, extracellular recordings in omPFC have revealed neurons that encode particular flavors in a manner that is modulated by satiety. That is, a neuron may "prefer" raisins to pieces of apple, but its responses decrease in intensity as the animal gets closer to

having eaten its fill of raisins. Thus, omPFC is a logical place to look for the representation of a common currency. Before we can start looking, though, we need to establish some definitions.

Transitivity, value transitivity, and menu invariance

A key property conferred by a common currency is *transitivity*. Let's return to my wine selecting dilemma from the beginning of the chapter, but simplify it, such that on one trip to my neighborhood wine shop (enigmatically named "Mike's," even though it's owned by Pat) there are only two varieties available, and I select bottle A over bottle B. Then, on a subsequent visit, the shop has sold out of A and has replaced it with C. Given this choice, I select B over C. On a third visit, if I find A and C on the shelf, I will choose A. (To do otherwise would be to show "erratic decision making.") Another property conferred by a common currency is that the value of one item can be expressed in terms of a quantity of any other, what economists call *value transitivity*. Thus, let's also say that my wine preferences are similar in proportion to those of the monkey whose juice preferences are illustrated in *Figure 15.8.C*. If Mike's started selling wine B in bottles holding 1.3 times the volume (but at the same price), I'd be equally happy with a 975 ml bottle of B as with a 750 ml bottle of A. And for the same reason, I'd be equally happy, for the same price, leaving the shop with four bottles of C as with one bottle of A. In economics, the term "menu" refers to a set of choices. Thus, the A vs. B choice is a menu, and the B vs. C choice is a different menu. For a neural implementation of a common currency to be effective, it needs to display "menu invariance," which means that, assuming equal levels of satiety, the level of activity representing B will be the same whether it is on the A-B menu or on the B-C menu. (One can imagine that the economist(s) who coined this naming convention may have done so over a meal, at a restaurant.)

Now we're ready for the next two courses on our menu.

Evidence from the nonhuman primate

Figure 15.8.D summarizes evidence, from a study by Padoa-Schioppa and Assad (2008), that orbitofrontal cortex (OFC) represents a common currency.

The neuron illustrated in the top row illustrates one type of menu invariance, in that its representation of the offer value for C is the same whether it's being offered against A or against B. The neuron illustrated in the middle row encodes the value of the choice that the animal made, regardless of which two juices were "on the menu." The neuron illustrated in the third row represents a different type of menu invariance, taste B, regardless of what else is "on the menu."

There are two details about the experimental procedure employed by Padoa-Schioppa and Assad (2008) that are worth highlighting before we move on. First, the spatial layout of offers was randomized across each experiment, such that the responses illustrated in *Figure 15.8.D* were invariant with regard to the retinotopic location of the representation of an offer and to the metrics of the saccade required to make a choice. In this way, the activity in OFC seems to differ from that in LIP, or in M1 and PMd, in that, for the latter, "value modulates responses that are sensory or motor in nature ... [whereas] neurons in the OFC encode economic value per se, independently of visuomotor contingencies" (Padoa-Schioppa and Assad, 2008; p. 101). The second detail is that, over the course of their experiment, they used 11 different juices, recombined into 23 different juice triplets. Thus, their findings are unlikely to be stimulus specific.

Why the OFC? In interpreting their findings, Padoa-Schioppa and Assad (2008) note that "Current anatomical maps divide the medial and orbital prefrontal cortices into 22 distinct brain areas, organized in two mostly segregated networks. In the orbital network (including 13m and the surrounding areas [*Figure 15.8.B*]), anatomical input from all sensory modalities converges with anatomical input from limbic areas including the amygdala. The orbital network thus seems well placed to compute a quantity such as subjective value" (p. 100) (*Tip 15.4*).

Still, *low-sugar tamarind Kool-Aid? Peppermint tea?* If these findings are to apply to human cognition, we wanna see 'em play out with factors that drive the behavior, and occupy the minds of modern humans, right? How about some money?!? How about some sex?!? (We'll get to drugs in an ensuing chapter. Not sure about the rock-and-roll, apart from my periodic gratuitous asides.)

FIGURE 15.8 The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Source: Padoa-Schioppa and Assad, 2008. Reproduced with permission of Nature Publishing Group.

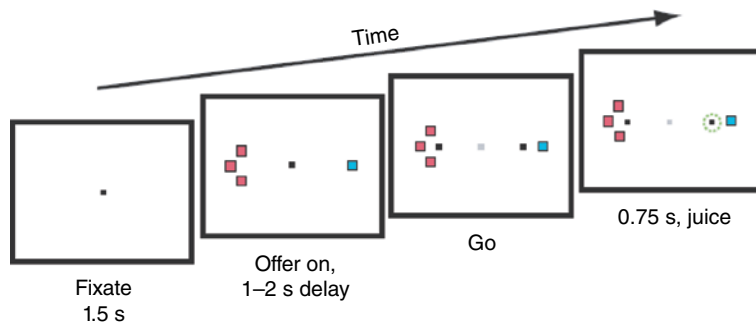


FIGURE 15.8.A On each trial, the monkey is made an offer that varies in terms of quantity and quality; in this case, three drops of agua frescas-flavored juice or one drop of peppermint tea. It indicated its choice by making a saccade to one of the two targets.

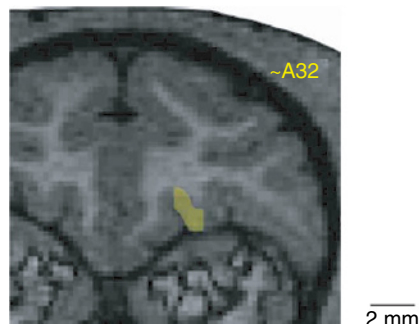


FIGURE 15.8.B Coronal slice of an MRI, with yellow shading indicating the location of area 13m from which recordings were made. Note the orbits of the eyes, immediately ventral to this portion of cortex.

A = grape, B = fruit punch, C = 1/2 apple

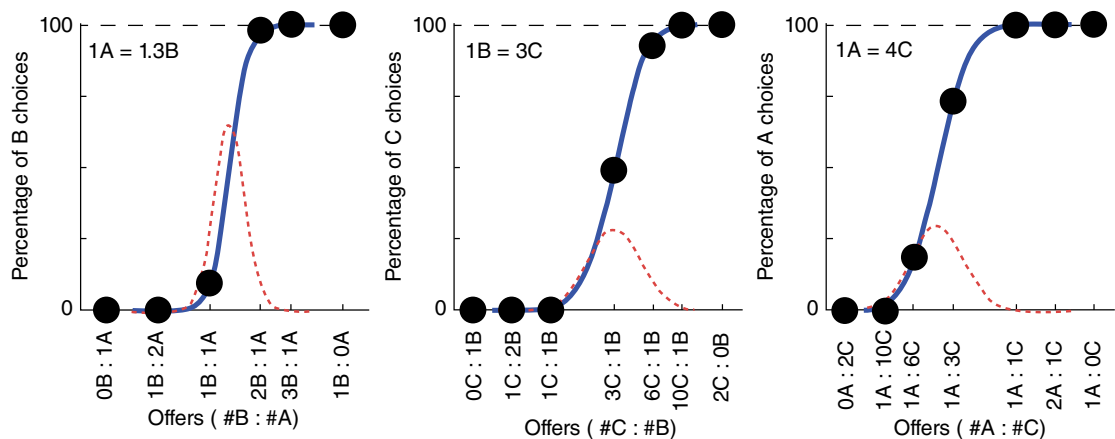


FIGURE 15.8.C Example of choice patterns from a session, featuring the three juices listed at top. This monkey liked grape juice better than fruit punch, and fruit punch better than apple juice diluted with equal parts water (thus, “1/2 apple”). Estimation from the sigmoid fit indicates that the monkey valued 1 unit of A the same as 1.3 units of B, 1 unit of B the same as 3 units of C, and one unit of A the same as 4 units of C. The dotted red lines illustrate the underlying normal distributions, their width indicating the variability in these estimates.

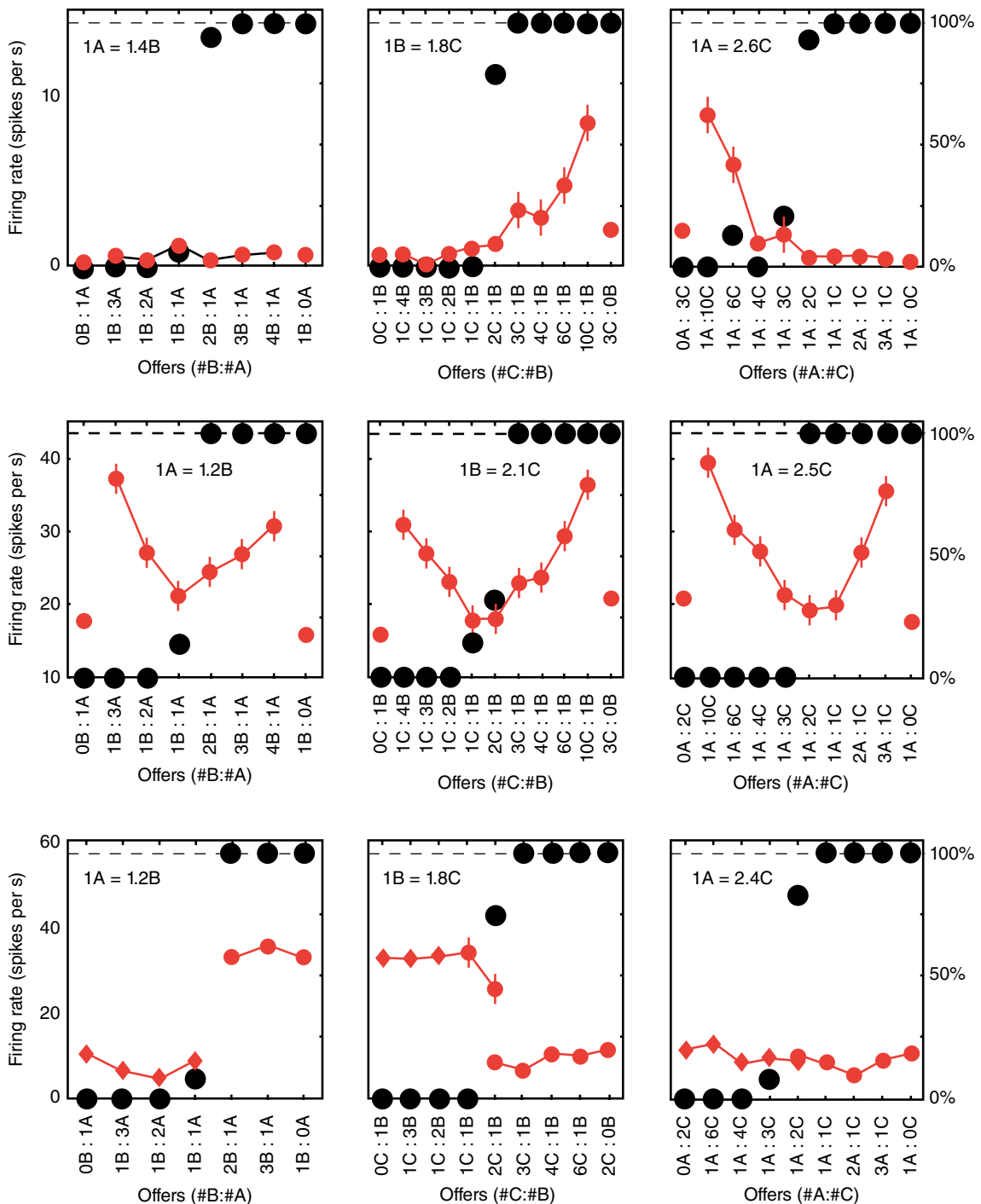


FIGURE 15.8.D Examples of three types of neurons in OFC. Each row of plots corresponds to a single neuron, with black dots indicating choices, with same conventions as *Figure 15.8.C*, except the ordinate for choices is on right side of each plot. Red dots represent firing rate. Neuron in top row encodes offer value C. Neuron in middle row encodes the "chosen value" (i.e., the difference in value between the chosen and unchosen offer, independent of the juice pair). Neuron in the bottom row encodes taste B, independent of pairing.

Evidence from humans

The study we'll next describe, by Smith et al. (2010), has been identified as the first to provide direct evidence for a common-currency representation of value in the human. It comes from a Duke University team that includes Michael Platt, whose group had previously performed a clever study of decision making with stimuli that were socially relevant for their male monkeys (Deaner, Khera, and Platt, 2005). Not unlike the Padoa-Schioppa and Assad (2008) study, it

Tip 15.4

Öngür and Price (2000) have delineated evidence for a medial network within omPFC that is heavily interconnected with the ventral striatum and sends efferents to visceromotor structures in the hypothalamus and brainstem. The orbital network, in contrast, receives heavy input from regions representing olfaction, gustation, visceral and somatic sensations, and vision, and its connections with more lateral regions of neostriatum suggest stronger integration with lateral regions of frontal and parietal cortex.

determined value transitivity in its subjects, doing so in a “pay per view” scheme in which male monkeys evaluated offers pitting juice against the opportunity to view images of the backsides of female monkeys in estrus, or the opportunity to view images of male monkeys from the same colony and with a known dominance status relative to the animal in the study. In a follow-up to the monkey pay-per-view study, Platt's group established (to no one's surprise) that heterosexual human males will trade small amounts of money or expend effort to view photographs of attractive females.

Building on this context, Smith and colleagues (2010) carried out a two-part study on 26 self-reported heterosexual males, ranging in age from 18 to 28. First, during fMRI, subjects passively viewed faces of young adult women varying in attractiveness from “very attractive” to “very unattractive,” and of images of paper money that varied in value from \$5 to –\$5 (*Figure 15.9.A*). (Attractiveness ratings were made by an independent group of raters on >2000 images, rumored to have been downloaded from www.hotornot.com.) The rationale for this first stage was to determine whether neural representations of decision values are computed even when no overt decision is being made. Immediately thereafter, while

FIGURE 15.9 Common currency for money and social reward in human vmPFC. Source: Smith, Hayden, Truong, Song, Platt, and Huettel, 2010. Reproduced with permission of the Society of Neuroscience.

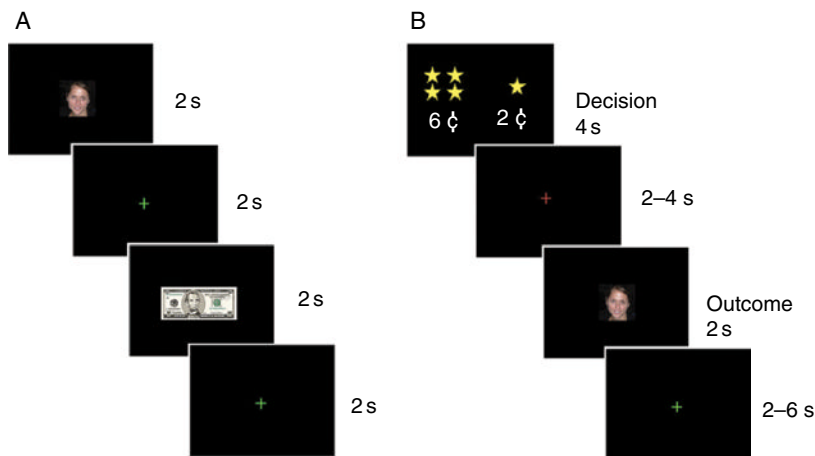


FIGURE 15.9.A AND B The first portion of the fMRI study **A** entailed no overt decision making, just passive viewing of images of faces that varied in attractiveness, and images of currency that varied in magnitude. The second portion **B** was an economic exchange task in which subjects chose, for each offer, whether to spend a larger sum of money or a smaller sum of money to view a more or less attractive face (indicated by the number of stars), respectively.

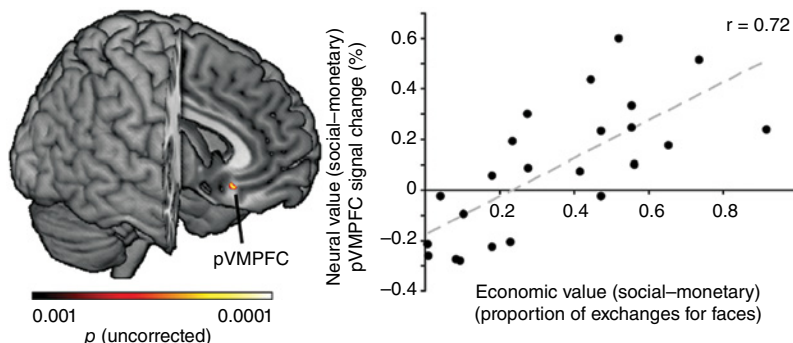


FIGURE 15.9.C Activity in posterior vmPFC, during passive viewing, predicts individual differences in willingness to exchange money to view attractive faces.

still in the scanner, subjects participated in an economic exchange task formally identical to that of Padua-Schioppa and Assad (2008), but instead of choosing between less volume of a more-preferred juice vs. more volume of a less-preferred juice, on each trial subjects chose between spending more money to view a “more-preferred” face vs. spending less money to view a “less-preferred” face. (In *Figure 15.9.B*, four stars corresponds to “most attractive,” one star to “least unattractive.”)

Results from the economic exchange task revealed considerable individual variability in each subject’s value transitivity scale – each person’s “willingness to pay” (WTP) to view a more attractive face ranged from 0 to 6.4 cents/star. The authors also computed a related measure, the proportion of offers on which the subject was willing to exchange more money to view a more-attractive face. This is the measure represented on the abscissa of the plot in *Figure 15.9.C*. On the ordinate is a measure of neural activity derived from the passive viewing task, computed to operationalize the extent to which brain activity “preferred” faces over money, or vice versa, with the contrast [(4-star-face – 1-star-face) – (large-monetary-gains – large-monetary-losses)]. As illustrated in *Figure 15.9.C*, the one region in the brain in which this neural measure of social value predicted individual differences in value transitivity scale was located in posterior vmPFC.

vmPFC or OFC?

What are we to make of the fact that the study of Smith et al. (2010) implicates human vmPFC as representing the common currency whereas that of

Padua-Schioppa and Assad (2008) implicates OFC in the macaque? It is certainly the case that the two-networks organizational scheme summarized in *Tip 15.4* is believed to apply to both species. As of this writing, this question remains unresolved, being glossed over, for example, in a 2012 meta-analysis of fMRI studies by Levy and Glimcher that, while by and large reinforcing the conclusion from the Smith et al. (2010) study, concludes that the neural seat of the common currency is “vmPFC/OFC.”

Has neuroeconomics taught us anything about the economics of decision making?

Almost every one of the papers that has been cited in this section on value-based decision making has cited decades-old principles and laws from behavioral economics: von Neumann and Morgenstern (1944); Samuelson (1937; 1947); Nash (1950a; 1950b); and so on. And so, cognitive neuroscientists, using the most cutting-edge neurophysiological measurement and analysis techniques available are, in effect, looking for evidence that 75-year-old ideas might be implemented in the brain? Certainly, none of the work that was reviewed here discovered any new laws of economics. To address this question, let’s return to the same review with which we kicked off this section:

we should be clear about the relationship between ... neurophysiological model[s] of choice and the very similar theoretical models in economics from which [they are] derived. Traditional economic models aim only to

predict (or explain) an individual's observable choices. They do not seek to explain the (putatively unobservable) process by which those choices are generated. In the famous terminology of Milton Friedman, traditional economic models are conceived of as being “as if” models (Friedman, 1953). Classic proofs in utility theory (i.e., Samuelson, 1937), for example, demonstrate that any decision maker who chooses in a mathematically consistent fashion behaves as if they had first constructed and stored a single list of the all possible options ordered from best to worst, and then in a second step had selected the highest ordered of those available options. Friedman and nearly all of the neoclassical economists who followed him were explicit that the concept of utility was not meant to apply to anything about [its possible instantiation in the brain]. For Friedman, and for many contemporary economists (i.e., Gul and Pesendorfer, 2008), whether or not there are “neurophysiological correlates of utility” is, by construction, irrelevant. (Kable and Glimcher, 2009, pp. 733–734)

The title of the Gul and Pesendorfer (2008) paper, for example, is “The Case for Mindless Economics.” (Get it?) Such skepticism toward the study of the brain bases of what one studies certainly isn't limited to economists. It was noted in the *Introduction to Section I* of this book, for example, that there has been a considerable amount of push-back against cognitive neuroscience from cognitive psychology (*Tip 15.5*).

But, as my 10-year-old daughter might say: “What' everrrr.”

First, just to state the obvious, unless our economist friends are dyed-in-the-wool dualists, they have to concede that these laws that do such an exquisite job of explaining human behavior must arise from the workings of the human brain. At a more pragmatic level, it's not their job to figure out how their laws are implemented by the brain. That's fine. Another branch of their discipline has ideas about how people who are

good at one thing should focus on that one thing, and let other people do other things. Those who study neuroeconomics are not trying to derive new laws of choice. They are seeking to understand the neural mechanisms of an aspect of human behavior that influences our every waking moment, one whose dysfunction leads to misery and, in some cases, tragedy, that plays out thousands of times every day. Addiction. Depression. Obesity. Psychopathy. These are just some of the health problems that neuroeconomics research may help us to better understand, and to palliate. (Finally, it should be noted that there are many formally trained economists who have gravitated to neuroeconomics as one way to pursue their scholarly interests; one such individual is featured at *Web Link 15.1*.)

And so, *Has neuroeconomics taught us anything about the economics of decision making?*

Wrong question.

FORAGING

And now for the third category of decision making invoked at the beginning of this chapter. There's a great deal of overlap between behavioral economics, psychological decision theory, and studies of group dynamics and optimal foraging behavior carried out under such disciplinary rubrics as anthropology, ecology, zoology, and dynamical systems. For example, the latter contribute importantly to many of the reviews by Glimcher and colleagues that we drew on in the previous section. Here, we'll consider one perspective that will give a feel for how this approach differs from the two that we have considered up to this point. Its consideration will also (once again) emphasize the overlap between the domains of “decision making” and “cognitive control.”

For much of his career, Oxford University-based cognitive neuroscientist Matthew Rushworth has explored the functions of the dorsomedial frontal cortex (dmFC), including the anterior cingulate cortex (ACC). Based on his training and many of his early studies, it's fair to say that he has approached this work from the perspective that this region is important for high-level motor control and response selection. (For context, see the discussion of Picard and Strick (1996)

Tip 15.5

It's rather arbitrary for this to come up in this chapter, but, nevertheless, an interesting and well-articulated debate about whether “neuroimaging has taught us anything about cognition” can be found in Coltheart (2006) and in Henson's (2006) reply.

in relation to *Figure 14.7*.) In recent work, Rushworth and colleagues have emphasized that whereas the majority of behavioral economics and neuroeconomics studies entail choices between two or more options, across evolutionary history primates have often been confronted with a different kind of decision. In particular, an organism often has to choose between remaining in the place where it is and exploiting the resources that are directly accessible, versus moving on to explore the environment for new, potentially more rewarding, opportunities. This foraging perspective, they argue, has implications for how one interprets activity in the dmFC.

In binary-choice decision-making tasks, dmFC signals are most often inversely related to the value difference between the chosen and unchosen option. This is what we saw in *Chapter 14 (Figure 14.7)*, in which difficult, or high-conflict, choices were associated with strong responses from dmFC. The alternative account from the foraging perspective, however, is that the error/difficulty/conflict accounts of this region's function may be artifacts of the two-alternative forced-choice tasks with which the majority of decision-making research is carried out. When assessed in foraging situations, dmFC signal strength is seen not to represent the unchosen option (and, therefore, a conflict), but instead to represent the relative benefit of exploring. One consequence of this view is that there may, in fact, be more than one common currency that guides decision making. Yes, by the foraging account, the research that we reviewed in the previous section that points to common currency representation in vmPFC/OFC is valid, and its conclusions generally solid. In addition, however, it argues that qualitatively different information may govern foraging decisions. In a foraging context there are three critical factors: the value of exploiting the option that one has encountered ("encounter value"); the potential value of exploring and finding something better ("search value"); and the energetic cost of embarking on the search ("search cost"). These three factors do not seem to be represented in the vmPFC/OFC, but are represented in the dmFC. Can a foraging-influenced model account for presumed functions of the dmFC that have, in other experimental contexts, been framed in terms of error, difficulty, and/or conflict? Developments in this literature will be among

the more exciting to watch over the next several years. (Indeed, during the few months separating when I wrote these words and when I corrected the final proofs for this book, a challenge to the foraging interpretation of dmFC has appeared, in the form of the *Further Reading* from Shenhav, Straccia, Cohen, and Botvinick (2014).

I CAN'T DECIDE!

Well, of course, that's not true, because you're reading the outcome of my decision. This is such a vibrant, dynamic area of research that several very difficult choices have had to have been made in order to limit this chapter to its prescribed length. I take some solace, however, in the fact that, not only did decision-making content leach into the previous chapter, it will occupy sizeable chunks of the next few as well. For now, we'll conclude this chapter with two boxes, one addressing a practical, social, and public-health application of decision making (*Real-World Applications 15.1*), the other addressing one of the deepest philosophical questions ever contemplated by a human being (*Thought Box 15.1*).

NEXT STOP

The South Shore of Long Island, a suburb of New York City, is so densely populated that what were once discrete villages have long-since merged into mile upon mile of, to an outsider's eye, unbroken suburban landscape. Although different towns have their own distinct character, without signs along the road (or announcement by the Long Island Rail Road conductor) one wouldn't know when one were leaving one town and entering another. The same might be said of *Chapter 15*, *Chapter 16*, and *Chapter 17* – each will have a different emphasis, but the transitions between them are arbitrary, and wouldn't be noticeable without explicit chapter breaks. Prominent in the landscape of each is the estimation and representation of value, the influence of social interactions, and interactions between medial frontal cortex and subcortical structures. So keep the facts and principles that we've just covered close at hand!

REAL-WORLD APPLICATIONS

15.1 Urchins! Cretins! Punks!

So I'm sitting out the patio in front of my house, writing a chapter about decision making. (True story.) I live in a leafy, residential neighborhood that is aptly named "University Heights." It's midmorning on a summer day and over the hill comes a group of six or so high-school-aged boys. Boisterous, jocular; you know, adolescent boys. They're obviously middle class, and in this neighborhood, the odds are that at least one of their mothers or fathers is on the faculty at the university where I teach. As they cross the street, walking through the neighbor's landscaping, one teases another sarcastically: "Oh nice, step on the flowers." Another retorts "Only step on the ones that are in bloom!" None of them actually do this, but when one notices me, he nonetheless shouts "Go Memorial," in a half-hearted attempt to fool me into thinking that they attend a different, rival high school, rather than the school in this neighborhood where they are almost certainly enrolled. The last one in the group makes eye contact me, then, running to catch up with his buddies, deliberately kicks an oak tree sapling growing in my side yard.

The stakes were low, the consequences minor, but those kids just made some bad decisions. Once I decided not to chase them down the street (*I didn't have shoes on ... they hadn't actually done anything wrong ... and what would I do once I caught up to them? ...*), I was reminded of a presentation I had recently attended by developmental cognitive neuroscientist B. J. Casey. The take-home: had any of these boys been walking down the street by himself, he very likely would not have walked through the bed of flowers, nor maliciously kicked my tree.

"Adolescence," as Casey, Jones, and Somerville (2011) have noted, "is a developmental period often characterized as a time of impulsive and risky choices leading to increased incidence of unintentional injuries and violence, alcohol and drug abuse, unintended pregnancy and sexually transmitted diseases" (p. 21). Casey's assessment, however, is that this is more of a social phenomenon than an indictment of the decision-making capabilities of any individual adolescent: it's not that their decision making is inherently bad, she argues, but that it's disproportionately susceptible to influence by what is popularly known as "peer pressure."

To illustrate this point, Casey's then-postdoctoral fellow Leah Somerville and colleagues (2013) scanned 69 subjects between the ages of 8 and 23. Subjects were misled into believing that there was a camera embedded in the MRI scanner, that during "video on" portions of the scan, their face would be viewed in real-time by a same-sex age-matched peer. In addition to fMRI, measures of self-reported embarrassment and skin conductance (i.e., sweatiness) were collected. Model fits indicated that peak embarrassment associated with "being watched" occurred at age 17.2, peak skin conductance at age 14.4, and peak activity in a medial PFC region associated with social cognition and emotional valuation processes at age 15.8. The authors propose that basal ganglia systems implicated in appetitive behavior are influenced by such social pressure, and that underdeveloped dorsal PFC regions that implement cognitive control can't counteract this susceptibility. *More at Web Clip 15.2.*

THOUGHT BOX

15.1 Free will?

Here's a quote that appears on the home page of the hayneslab Web site:

Decisions don't come from nowhere but they emerge from prior brain activity. Where else should they come from? In theory it could be possible to trace the causal pathway of a decision all the way back to the big bang. Our research shows that we can trace it back 10 seconds. Compared to the time since the big bang this is not very long.

Stick *that* in your pipe and smoke it.

Berlin-based John-Dylan Haynes has been a pioneering, influential contributor to the multivariate pattern analysis (MVPA) of fMRI data. The study described here, by Soon, Brass, Heinze, and Haynes (2008), may be the one for which he is best known, certainly in the popular media. In it, subjects viewed a 2 Hz serial presentation of individual letters (*Figure 15.10.A*) and, periodically, “when they felt the urge to do so” (p. 543), were to press a button with the left or right index finger. After each button press, a display appeared on which they indicated their best estimate of the letter that was on the screen at the time that they decided to move. Button presses occurred an average of 21.6 seconds after the onset of each trial.

FIGURE 15.10 Unconscious determinants of free decisions in the human brain. Source: Soon, Brass, Heinze, and Haynes, 2008. Reproduced with permission of Nature Publishing Group.

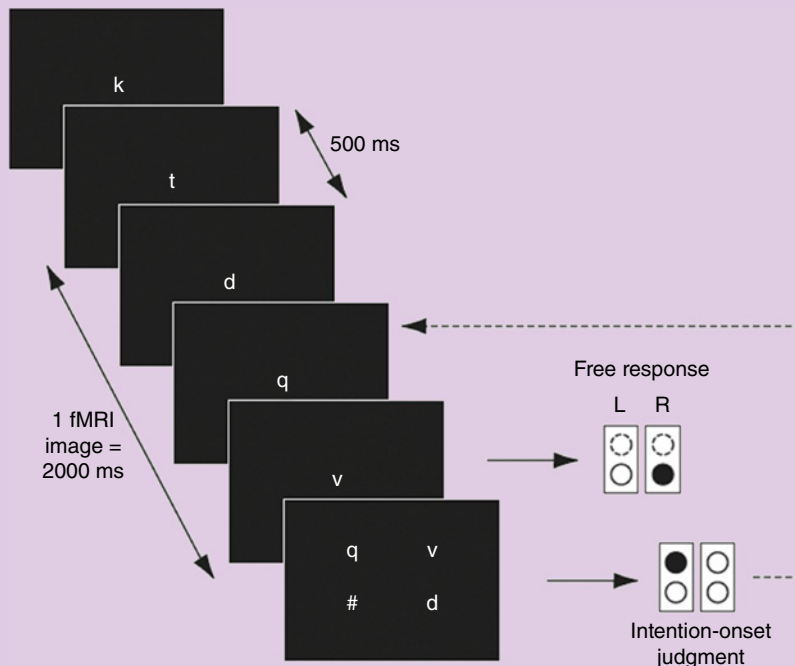


FIGURE 15.10.A Illustration of a trial on which the subject pushed the right-hand button while the letter “v” was on the screen. Immediately thereafter s/he indicated which of the following had been on the screen when s/he made the decision to press the button: “v,” “q,” “d,” or a previous letter (“#”). On nearly 90% of trials, subjects indicated that the decision had been made within 1 second of when the choice was executed.

Thought Box 15.1 *Continued*

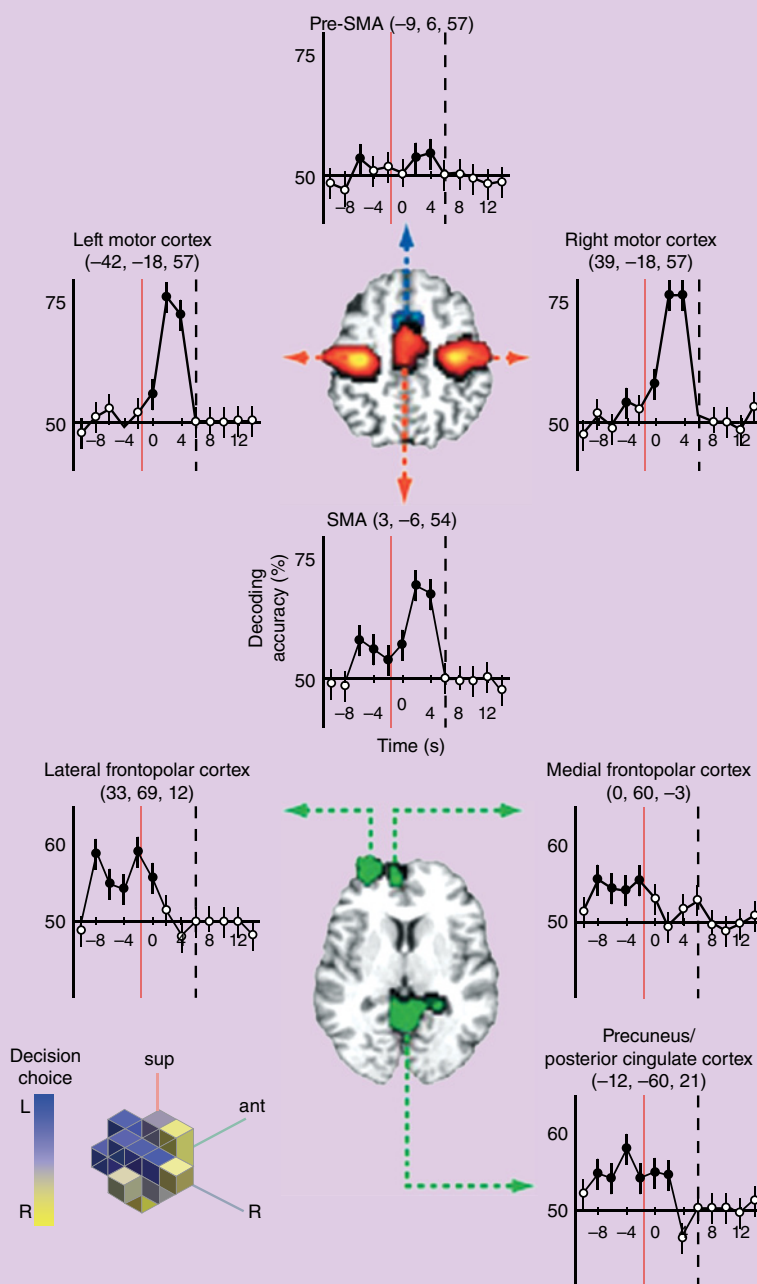


FIGURE 15.10.B Results of decoding the finger to be used for the impending, volitionally initiated and executed, button press. Each plot shows decoding performance for an ROI, time-locked to the button press (time 0). For each, the vertical red line indicates the average time at which the subjects reported becoming aware of the decision to press the button; the vertical dashed line marks the beginning of the next trial. Color coding of ROIs distinguishes those from which button choice could be predicted prior to reported awareness (green) from those from which it could be predicted only after this time (red).

Thought Box 15.1 Continued

The data were analyzed with a “roving searchlight” application of MVPA, in which, at a given voxel in the brain, a spherical cluster with radius of three voxels was defined, and MVPA applied to just the voxels in that sphere. Then the sphere was moved over by one voxel and the process repeated, for every gray-matter voxel in the brain. The lower-left of *Figure 15.10.B* illustrates one such sphere, from frontopolar cortex, color-coded to illustrate how its activity distinguished whether the next button presses would be with the left or right finger.

The provocative finding from this study, illustrated in *Figure 15.10.B*, is that, reliably across several subjects, two regions of the brain were shown to contain information about the next button press a full 10 seconds before subjects reported having become aware of deciding to make that press! It’s as though “their brains knew what they were going to do before they, themselves, did”!?! Neither of the two regions – located in the frontopolar and medial parietal/posterior cingulate cortex – is directly implicated in motor control. Regions that are, the SMA and M1, didn’t contain side-of-button-press information until the time at which the event was happening.

Interestingly, however, in a separate analysis, the authors applied MVPA to determine which regions of the brain predicted the *time* at which each button press would occur (regardless of which button). This analysis revealed that the lateral and midline motor regions represented this information earlier than did the frontopolar and medial parietal areas. Indeed, the ramping-up shape of the decoding time courses in SMA and pre-SMA – beginning 6 to 8 seconds prior to the button press – are reminiscent (although on a different time scale) of the build-up of accumulated evidence seen in LIP during RDM decision tasks requiring an oculomotor response (*Figure 15.3*) and in motor areas during RDM decision tasks requiring a skeletomotor response (*Figure 15.6*). (*Web Clip 15.3* provides a more detailed treatment of this study.)

END-OF-CHAPTER QUESTIONS

1. What are three different types of decisions that we need to make, and what distinguishes each?
2. How has microstimulation informed our understanding of perceptual decision making about the direction of motion in RDM displays?
3. How does the behavior of neurons in LIP relate to drift-diffusion models of perceptual decision making?
4. Why is the activity of LIP neurons so different on two-alternative forced-choice trials, requiring identical saccades, when the task is a RDM perceptual decision task vs. when it is a free choice between different levels of expected gain?
5. How do the concepts of transitivity, value transitivity, and menu invariance relate to the idea of a common currency? In what brain area are all these properties observed?
6. Has neuroeconomics taught us anything about the economics of decision making?
7. How do foraging decisions differ fundamentally from perceptual and value-based decisions? What brain area is implicated in the control of foraging?
8. Do we have free will over the control of our behavior?

REFERENCES

- Casey, B. J., Rebecca M. Jones, and Leah H. Somerville. 2011. “Braking and Accelerating of the Adolescent Brain.” *Journal of Research on Adolescence* 21 (1): 21–33. doi: 10.1111/j.1532-7795.2010.00712.x.

- Coltheart, Max. 2006. "What Has Functional Neuroimaging Told Us about the Mind (So Far)?" *Cortex* 42 (3): 323–331. doi: 10.1016/S0010-9452(08)70358-7.
- Deaner, Robert O., Amit V. Khera, and Michael L. Platt. 2005. "Monkeys Pay Per View: Adaptive Valuation of Social Images by Rhesus Macaques." *Current Biology* 15 (6): 543–548. doi: 10.1016/j.cub.2005.01.044.
- Donner, Tobias H., Markus Siegel, Pascal Fries, and Andreas K. Engel. 2009. "Buildup of Choice-Predictive Activity in Human Motor Cortex during Perceptual Decision Making." *Current Biology* 19 (18): 1581–1585. doi: 10.1016/j.cub.2009.07.066.
- Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Glimcher, Paul W. 2002. "Decisions, Decisions, Decisions: Choosing a Biological Science of Choice." *Neuron* 36 (2): 323–332. doi: 10.1016/S0896-6273(02)00962-5.
- Gold, Joshua I., and Michael N. Shadlen. 2007. "The Neural Basis of Decision Making." *Annual Review of Neuroscience* 30: 535–574. doi: 10.1146/annurev.neuro.29.051605.113038.
- Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." In *The Foundations of Positive and Normative Economics: A Handbook*, edited by Andrew Caplin and Andrew Schotter, 3–41. New York: Oxford University Press.
- Hanks, Timothy D., Jochen Ditterich, and Michael N. Shadlen. 2006. "Microstimulation of Macaque Area LIP Affects Decision-Making in a Motion Discrimination Task." *Nature Neuroscience* 9 (5): 682–689. doi: 10.1038/nn1683.
- Henson, Richard. 2006. "What Has (Neuro)Psychology Told Us about the Mind (So Far)? A Reply to Coltheart (2006)." *Cortex* 42 (3): 387–392. doi: 10.1016/S0010-9452(08)70365-4.
- Kable, Joseph W., and Paul W. Glimcher. 2009. "The Neurobiology of Decision: Consensus and Controversy." *Neuron* 63 (6): 733–745. doi: 10.1016/j.neuron.2009.09.003.
- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–291. doi: 10.2307/1914185.
- Levy, Dino J., and Paul W. Glimcher. 2012. "The Root of All Value: A Neural Common Currency for Choice." *Current Opinion in Neurobiology* 22 (6): 1027–1038. doi: 10.1016/j.conb.2012.06.001.
- Lo, Chung-Chuan, and Xiao-Jing Wang. 2006. "Cortico-Basal Ganglia Circuit Mechanism for a Decision Threshold in Reaction Time Tasks." *Nature Neuroscience* 9 (7): 956–963. doi: 10.1038/nn1722.
- Nash, John R. 1950a. "The Bargaining Problem." *Econometrica: Journal of the Econometric Society* 18: 155–162.
- Nash, John R. 1950b. "Equilibrium Points in n-Person Games." *Proceedings of the National Academy of Sciences of the United States of America* 36: 48–49.
- Newsome, William T., Kenneth H. Britten, and Anthony Movshon. 1989. "Neuronal Correlates of a Perceptual Decision." *Nature* 341 (6237): 52–54. doi: 10.1038/341052a0.
- Öngür, Dost, and Joel L. Price. 2000. "The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans." *Cerebral Cortex* 10 (3): 206–219. doi: 10.1093/cercor/10.3.206.
- Padoa-Schioppa, Camillo, and John A. Assad. 2008. "The Representation of Economic Value in the Orbitofrontal Cortex Is Invariant for Changes of Menu." *Nature Neuroscience* 11 (1): 95–102. doi: 10.1038/nn2020.
- Picard, Nathalie, and Peter L. Strick. 1996. "Motor Areas of The Medial Wall: A Review of their Location and Functional Activation." *Cerebral Cortex* 6: 342–353. doi: 10.1093/cercor/6.3.342.
- Platt, Michael L., and Paul W. Glimcher. 1999. "Neural Correlates of Decision Variables in Parietal Cortex." *Nature* 400 (6741): 233–238. doi: 10.1038/22268.
- Salzman, C. Daniel, Kenneth H. Britten, and William T. Newsome. 1990. "Cortical Microstimulation Influences Perceptual Judgements of Motion Direction." *Nature* 346 (6280): 174–177. doi: 10.1038/346174a0.
- Samuelson, Paul A. 1937. "A Note on the Measurement of Utility." *Review of Economic Studies* 4 (2): 155–161. doi: 10.2307/2967612.
- Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Vol. 80 of *Harvard Economic Studies*. Cambridge, MA: Harvard University Press.
- Smith, David V., Benjamin Y. Hayden, Trong-Kha Truong, Allen W. Song, Michael L. Platt, and Scott A. Huettel. 2010. "Distinct Value Signals in Anterior and Posterior Ventromedial Prefrontal Cortex." *Journal of Neuroscience* 30 (7): 2490–2495. doi: 10.1523/JNEUROSCI.3319-09.2010.
- Somerville, Leah H., Rebecca M. Jones, Erika J. Ruberry, Jonathan P. Dyke, Gary Glover, and B. J. Casey. 2013. "The Medial Prefrontal Cortex and the Emergence of Self-Conscious Emotion in Adolescence." *Psychological Science* 24(8): 1554–1562. doi: 10.1177/0956797613475633.
- Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. 2008. "Unconscious Determinants of Free Decisions in the Human Brain." *Nature Neuroscience* 11 (5): 543–545. doi: 10.1038/nn.2112.
- Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

OTHER SOURCES USED

- Gold, Joshua I., and Michael N. Shadlen. 2000. "Representation of a Perceptual Decision in Developing Oculomotor Commands." *Nature* 404 (6776): 390–394. doi: 10.1038/35006062.
- Gold, Joshua I., and Michael N. Shadlen. 2001. "Neural Computations that Underlie Decisions about Sensory Stimuli." *Trends in Cognitive Sciences* 5 (1): 10–16. doi: 10.1016/S1364-6613(00)01567-9.
- Kolling, Nils, Timothy E. J. Behrens, Rogier B. Mars, and Matthew F. S. Rushworth. 2012. "Neural Mechanisms of Foraging." *Science* 336 (6077): 95–98. doi: 10.1126/science.1216930.
- Padoa-Schioppa, Camillo, and John A. Assad. 2006. "Neurons in the Orbitofrontal Cortex Encode Economic Value." *Nature* 441 (7090): 223–226. doi: 10.1038/nature04676.
- Ratcliff, Roger, Anil Cherian, and Mark Segraves. 2003. "A Comparison of Macaque Behavior and Superior Colliculus Neuronal Activity to Predictions from Models of Two-Choice Decisions." *Journal of Neurophysiology* 90 (3): 1392–1407. doi: 10.1152/jn.01049.2002.
- Ratcliff, Roger, and Jeffrey N. Rouder. 1998. "Modeling Response Times for Two-Choice Decisions." *Psychological Science* 9 (5): 347–356. doi: 10.1111/1467-9280.00067.
- Rushworth, Matthew F. S., Nils Kolling, Jérôme Sallet, and Rogier B. Mars. 2012. "Valuation and Decision-Making in Frontal Cortex: One or Many Serial or Parallel Systems?" *Current Opinion in Neurobiology* 22 (6): 946–955. doi: 10.1016/j.conb.2012.04.011.
- Shadlen, Michael N., and William T. Newsome. 1996. "Motion Perception: Seeing and Deciding." *Proceedings of the National Academy of Sciences of the United States of America* 93 (2): 628–633. doi: 10.1073/pnas.93.2.628.
- Shadlen, Michael N., and William T. Newsome. 2001. "Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey." *Journal of Neurophysiology* 86 (4): 1916–1936.

FURTHER READING

- Glimcher, Paul W. 2010. *Foundations of Neuroeconomic Analysis*. Oxford: Oxford University Press.
As of this writing, the definitive statement of the development, current status, and future potential of neuroeconomics.
- Gold, Joshua I., and Michael N. Shadlen. 2007. "The Neural Basis of Decision Making." *Annual Review of Neuroscience* 30: 535–574. doi: 10.1146/annurev.neuro.29.051605.113038.
Authoritative review of visual perceptual decision making, including thorough treatment of application of SDT.
- Kable, Joseph W., and Paul W. Glimcher. 2009. "The Neurobiology of Decision: Consensus and Controversy." *Neuron* 63 (6): 733–745. doi: 10.1016/j.neuron.2009.09.003.
A broad-ranging and synthetic overview of neuroeconomics.
- Rushworth, Matthew F. S., Nils Kolling, Jérôme Sallet, and Rogier B. Mars. 2012. "Valuation and Decision-Making in Frontal Cortex: One or Many Serial or Parallel Systems?" *Current Opinion in Neurobiology* 22 (6): 946–955. doi: 10.1016/j.conb.2012.04.011.
A comprehensive summary of the foraging perspective of the role of dmFC in control and decision making.
- Schultz, Wolfram. 2013. "Updating Dopamine Reward Signals." *Current Opinion in Neurobiology* 23 (2): 229–238. doi: 10.1016/j.conb.2012.11.012.
Introduced in the previous chapter, but central to decision making, the most up-to-date statement, at the time of this writing, of our understanding of dopamine reward signals.
- Shenhav, Amitai, Mark A. Straccia, Jonathan D. Cohen, and Matthew M. Botvinick. 2014. "Anterior Cingulate Engagement in a Foraging Context Reflects Choice Difficulty, Not Foraging Value." *Nature Neuroscience* 17: 1249–1254. doi: 10.1038/nn.3771.
A response to the foraging interpretation of cognitive control-related functions of dmFC.

CHAPTER 16

SOCIAL BEHAVIOR

KEY THEMES

- The trustworthiness of an actor in an economic exchange influences one's delay-discounting function – a variable from neuroeconomics – highlighting the fact that social context can influence many, if not all, of the behavioral and cognitive processes that we have studied to date.
- The case of Phineas Gage, whose medial PFC was destroyed in a freak rock-blasting accident, and whose personality was dramatically altered as a consequence, highlights the role of this region in controlling social behavior.
- Neural abnormalities in the vmPFC and its connections to medial temporal and inferior parietal regions may be a factor in the “psychiatric” condition of psychopathy.
- Neuroscientific study of “theory of mind” (ToM), the ability to represent the thoughts and motivations of others, is an informative lens through which to observe how scientists go about studying the neural bases of social behavior – the methods, the theoretical perspectives, the complications, and the controversies.
- One hotly contested conceptual battleground has been whether the functions of the right temporoparietal junction are better characterized as being specific to ToM and moral decision making, or more generally recruited by tasks requiring attentional reorienting and, perhaps, the comparison of internal models of the world with external sensory evidence.
- Observing the behavior of others – an inherently social act – can be an important way to learn new information, even if we only observe others' actions but not the outcome(s) of those actions. Model-based analysis of fMRI provides evidence that different kinds of error prediction signals, processed by different brain areas at different times during learning, may underlie this otherwise perplexing phenomenon.

CONTENTS

KEY THEMES

TRUSTWORTHINESS: A PREAMBLE

Delaying gratification: a social influence on a
“frontal” class of behaviors

THE ROLE OF vmPFC IN THE CONTROL OF SOCIAL COGNITION

Phineas Gage

Horrible accident

Sequelae

Contemporary behavioral neurology

*Linking neurological and psychiatric
approaches to social behavior*

THEORY OF MIND

The ToM network

Two asides about anterior paracingulate cortex

The temporoparietal junction (TPJ)

Challenge to the specificity of TPJ for ToM

False beliefs (?) about Rebecca Saxe's mind

fMRI with higher spatial resolution

rTMS of TPJ during moral judgments

*Moral judgments, TPJ, and autism spectrum
disorder*

A final assessment of the role of TPJ in ToM
mentalization

OBSERVATIONAL LEARNING

TRUSTWORTHINESS, REVISITED

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

TRUSTWORTHINESS: A PREAMBLE

As highlighted by the anecdote in *Real-World Applications 15.1*, people often behave differently when others are present than when they are alone. Because all behavior derives from the workings of the CNS, it must be that such social influences on our behavior are, themselves, also mediated by the brain. This should be reason enough to want to study and understand them. But there's more to it than that. This brief preamble illustrates just one example of how those who wish to assemble a comprehensive model of the neural bases of human cognition need to incorporate social factors. Otherwise, at best, theirs will be an incomplete model.

Delaying gratification: a social influence on a “frontal” class of behaviors

We've seen many examples over the past few chapters, of failures of cognitive control being associated with the integrity of the PFC. One hallmark class of behaviors that we haven't considered explicitly is delaying gratification – being able to withhold the impulse to take an immediate reward in favor of achieving a longer-term goal. One's ability to delay gratification seems to be an important trait that predicts success in many aspects of life. For example, during the pre-school years, a period when the PFC is underdeveloped relative to other neural systems, it has been shown that the ability to resist a desirable immediate treat in favor of a larger delayed one predicts higher levels of success later in life: higher SAT scores and better social competence in adolescence; lower rates of obesity and of substance abuse in adulthood.

In behavioral economics, and, therefore, more recently, in neuroeconomics, delaying gratification has been operationalized with **intertemporal choice** tasks. These measure the phenomenon of **delay discounting**, demonstrated by the fact that people tend to prefer a smaller amount of money (or some other reward) immediately over the promise of a larger amount of money at a later date. (It's “as if” they are discounting the value of the larger amount, because of the delay.) Intertemporal choice behavior has been used by many

to index behavioral impulsivity, a factor that is sensitive to the integrity of frontostriatal systems. Further, because delay discounting is also observed in nonhuman animals, including rodents and monkeys, it is at a rich nexus of overlapping research domains, including behavioral neuroscience, behavioral neuropharmacology, behavioral genetics, and neuroeconomics. To list just a few examples, in rats, it can be studied via manipulation of neurotransmitter systems, such as serotonin, and in combination with drugs, like methylphenidate, that act on noradrenergic and dopaminergic systems; in the monkey, in relation to the activity of midbrain dopaminergic neurons; in humans, as a function of variation in genotype that relates to dopamine processing; and so on. It is seen as an important tool for studying factors underlying **attention deficit hyperactivity disorder** (ADHD).

What does this all have to do with social behavior? The team of Michaelson, de la Vega, Chatham, and Munakata (2013) have demonstrated an answer to this question by pointing out that social cognition adds another dimension to the phenomenon:

delaying gratification also relies on the fundamental assumption that a future reward will be delivered as promised ... [and so] is not simply about choosing “more later” over “some now,” but rather, requires choosing “*maybe* more later” over “some now,” when there is doubt about whether those promising the future reward would come through [i.e., would honor their commitment to deliver the reward in the future]. (p. 1)

To show this, they designed a study in which, on an intertemporal choice task, they manipulated the perceived trustworthiness of one party in the exchange. That's right, “trustworthiness.” (Yes, this is Yuko Munakata from *Research Spotlight 14.1*; so one might think that she'd be content to stick with studying the role of PFC in delaying gratification. But no ...)

Michaelson, de la Vega, et al. (2013) administered an intertemporal choice task in which adult subjects read three vignettes, depicting a person who was trustworthy, untrustworthy, or neutral, then judged delay gratification situations involving these individuals. (For example, “If [individual x] offered you \$40 now or \$65 in 70 days, which would you choose?”) Their findings show that the probability that one will delay gratification depends, in part, on the perceived trustworthiness

of the other actor in the exchange. And because one of their studies manipulated perceived trustworthiness as a within-subject factor, the influence of this factor must be, at least in part, independent of frontal-lobe functioning, neurotransmitter levels, genotype, and other explicitly neurobiological factors that are also known to influence intertemporal choice.

Now, let's think all the way back to *Thought Box 1.1*. The present preamble directly invokes the admonition that not all phenomena that have a psychological reality ("conscientiousness" in *Thought Box 1.1*, "trustworthiness" here) are necessarily amenable to direct study with the methods of cognitive neuroscience. That still seems reasonable to me. But now we find ourselves in a situation where, if we want to fully understand this aspect of behavior – delay discounting – we will have to start thinking about incorporating the neural processing of information about the trustworthiness of other parties in the exchange. Just walking away from the problem because "it would be very difficult to study with the methods of cognitive neuroscience" isn't an option. As noted previously, for example, this line of research on delay discounting has made important contributions to understanding ADHD. And so what this means is that a problem that we're interested in has (yet again) gotten more complicated.

Just a moment's reflection indicates that social factors are likely to influence many, if not all, of the behaviors and cognitive processes whose neural bases we have considered up to this point in this book. (For example, action selection/motor control? The word "social" didn't appear once in *Chapter 7*, yet the influence of social context on the skeletomotor control of that tree-kicking cretin called out in *Real-World Applications 15.1* is undeniable.) Thus, we need to develop an understanding of how to study aspects of behavior, and influences on behavior, that are inherently social.

THE ROLE OF vmPFC IN THE CONTROL OF SOCIAL COGNITION

We'll start our consideration of social neuroscience with this brain region that featured prominently in the previous chapter's treatment of value-based decision making.

Phineas Gage

If there's a neurological patient who might challenge H. M.'s status as most cited case in the brain sciences, it's Phineas Gage. The fantastical nature of his accident and injury, and its profound behavioral consequences, ensure that he appears in every textbook, and in more than his fair share of introductory slides in scientific presentations.

Horrible accident

In 1848, Gage was the foreman of a crew that was blasting rock to create a route through the mountains of Southcentral Vermont for a new line for the Rutland and Burlington Railroad. At that time, this entailed drilling a deep, narrow hole into the rock, filling it with blasting powder, then a fuse, then sand, compressing the charge with a metal tamping iron (of, say, 3'7'' in length and 1.25'' in diameter; and did we say fashioned out of *iron*?), then presumably walking a long distance away, yelling something like "Fire in the hole!" before lighting the fuse. An article from the *Boston Post* picks up the story from here:

Horrible Accident – As Phineas P. Gage, a foreman on the railroad in Cavendish, was yesterday engaged in tamping for a blast, the power exploded, carrying an iron instrument through his head an inch and a fourth in circumference [*sic*] and three feet and eight inches [*sic*] in length, which he was using at the time. The iron entered on the side of this face, shattering the upper jaw [*sic*], and passing back of the left eye, and out at the top of the head.

The most singular circumstance connected with the melancholy affair is, that he was alive at two o'clock this afternoon [i.e., on the following day], and in full possession of his reason, and free from pain – Ludlow, Vt., Union. (September 21, 1848)

(Everything qualified with a "*sic*" is not strictly accurate; but what's an inch here or there?) Most singular indeed. And the fact that he was to live for not just a day after the accident, but for nearly 12 additional years, and that a detailed assessment was made of the lasting consequences of his injuries, means that contemporary scientists have gone to great lengths to reconstruct the damage likely incurred by his brain (*Figure 16.1* and *Figure 16.2*). Details about how one group has gone about this are presented in *Web Clip 16.1*.



FIGURE 16.1 Daguerrotype of Phineas Gage, taken years after his accident, holding the tamping iron that was “his constant companion for the remainder of his life.” Source: Harlow, 1868.

Sequelae

Why all the fuss? Here’s the most-quoted section from the first thorough account in the medical literature of the behavioral consequences of Gage’s injury, published eight years after his death:

The equilibrium or balance, so to speak, between his intellectual faculties and animal propensities, seems to have been destroyed. He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operations, which are no sooner arranged than they are abandoned in turn for others appearing more feasible. A child in his intellectual capacity and manifestations, he has the animal passions of a strong man. Previous to his injury, although untrained in the schools, he possessed a well-balanced mind, and was looked upon by those who knew him as a shrewd, smart businessman, very

Tip 16.1

In 1877, David Ferrier (see *Chapter 1*) wrote the chair of Harvard’s department of psychology to inquire about reports he had heard about this case, noting, “In investigating reports on diseases and injuries of the brain, I am constantly amazed at the inexactitude and distortion to which they are subject by men who have some pet theory to support. The facts suffer so frightfully that I feel obliged always to go to the fountainhead – dirty and muddy though this frequently turns out” (Macmillan, 2000, p. 464).

energetic and persistent in executing all his plans of operation. In this regard his mind was radically changed, so decidedly that his friends and acquaintances said he was ‘no longer Gage.’ (Harlow, 1868, pp. 339–340)

The extent to which North American scientific culture was underdeveloped at that time, relative to that of Europe, is evident in the fact that it was decades before the scientific community (a) first learned of this case, then (b) fully accepted the veracity of its details (*Tip 16.1*), and finally (c) incorporated it into models of the behavioral functions of the medial PFC. (*Web Link 16.1* connects to more resources about this fascinating case.)

Contemporary behavioral neurology

Mesulam (1985) has summarized the frontal cortex as being organized into three divisions: motor-premotor, paralimbic, and heteromodal (*Tip 16.2*). It is the paralimbic that has been most strongly associated with social behavior. Comprising anterior cingulate and caudal orbitofrontal regions, it is characterized by a

Tip 16.2

What Mesulam (1986) refers to here as “paralimbic” and “heteromodal” PFC correspond, in cytoarchitectonic terms, to “agranular” and “granular” PFC, respectively. “Granular” refers to cortex having a large, conspicuous layer IV.

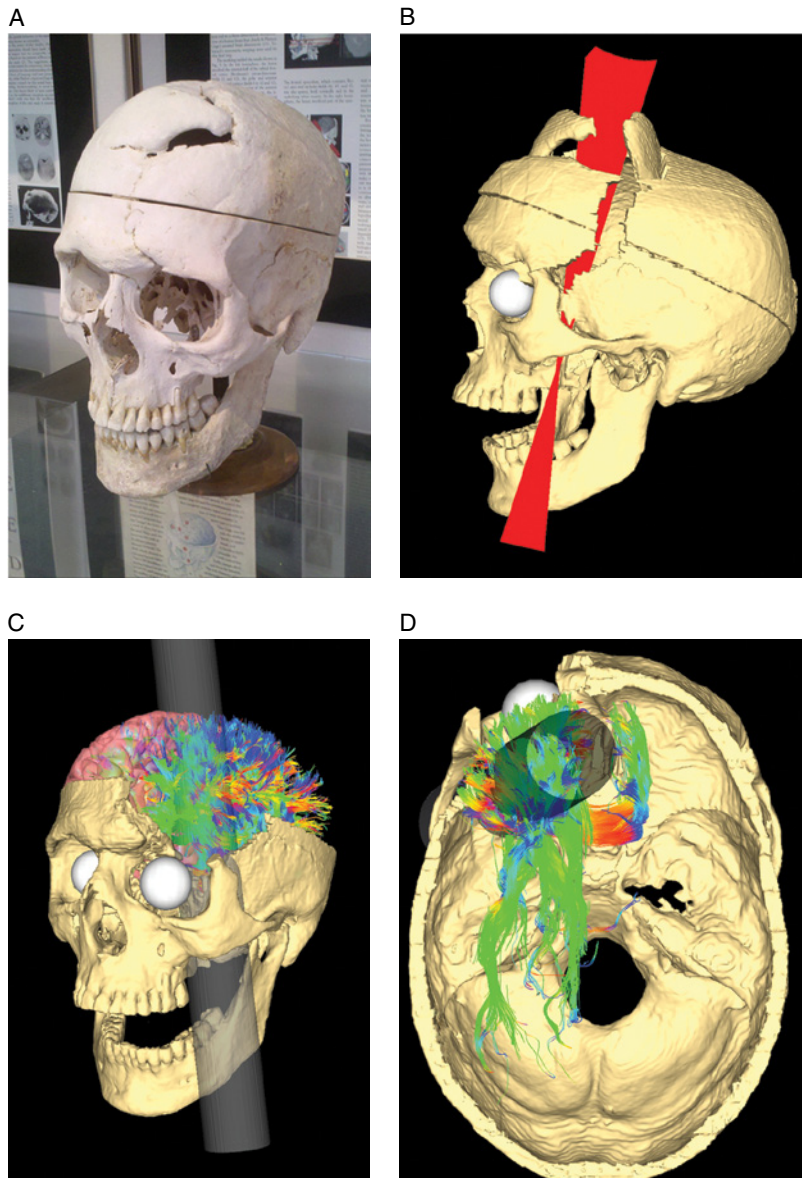


FIGURE 16.2 Reconstruction of damage to Phineas Gage's brain. **A.** The skull of Phineas Gage on display at the Warren Anatomical Museum at Harvard Medical School. **B.** The set of possible rod trajectory centroids, estimated from CT scans of the skull. **C.** A rendering of the Gage skull with the best fit rod trajectory and example fiber pathways in the left hemisphere intersected by the rod. **D.** Same data as C, but rendered in radiological conventions, and portrayed as though viewed from the underside, with a reconstruction of the ventral surface of the brain. Gray shading is estimate of frank lesion to the left frontal lobe (note that circle within this region indicates the actual diameter of the tamping iron). Fiber pathways estimated to have been severed by the path of the tamping iron are illustrated with colors representing the orientation of the tract (e.g., green is parallel to the Y dimension, yellow-red to X, and blue to Z). Source: Van Horn, Irimia, Torgerson, Chambers, Kikinis, and Toga, 2012. Reproduced with permission of the authors.

cytoarchitecture that is transitional from evolutionarily older archicortex of the hippocampus and piriform cortex of the olfactory system and heteromodal neocortex. (Note that the caudal orbitofrontal component of paralimbic cortex corresponds to the ventromedial (vm)PFC that featured prominently in the previous chapter.) In an editorial that accompanied the publication of the two reports from Lhermitte (1986) that we considered in *Chapter 14*, Mesulam (1986) notes that input from regions processing internal physiological states and “extensively preprocessed sensory information ... suggests that [paralimbic] areas should be of crucial importance for channeling drive and emotion to appropriate targets in the environment...” Then, with explicit reference to social behavior, he continues:

Experimental lesions in the paralimbic component of the frontal lobe also interfere with social interactions. In social animals such as monkeys, group bonds heavily rely on specific aggressive and submissive displays, grooming behavior, and vocalizations. Success depends on directing the proper behavior to the proper individual in the proper context. Animals with orbitofrontal lesions show a severe disruption of these conspecific affective and affiliative behaviors and eventually experience social isolation ... These experiments provide a model for the socially maladaptive and emotionally inappropriate behaviors seen after frontal lobe damage and support the contention that these behaviors emerge after involvement of the paralimbic component in the frontal lobe. (pp. 322–323)

Linking neurological and psychiatric approaches to social behavior

Consistent with the perspective represented in the preceding quote is work from Michael Koenigs and his colleagues, who have studied both patients with lesions of the vmPFC, and incarcerated inmates who have been diagnosed with psychopathy. Psychopathy is a psychiatric condition associated with callous and impulsive antisocial behavior, and the rationale behind some of Koenigs’ research is that “For decades, neurologists have noted that the personality changes accompanying vmPFC damage (e.g., lack of empathy, irresponsibility, and poor decision making) bear striking resemblance to hallmark psychopathic personality traits,” to the extent that “the personality changes associated with vmPFC damage have been dubbed ‘pseudopsychopathy’ and ‘acquired sociopathy’”

Tip 16.3

What distinguishes a “neurological” diagnosis from a “psychiatric” one? Mostly tradition within how medicine is/used to be practiced. “Back in the day,” one could get away with the half-joke that neurology concerns itself with conditions for which the neural bases are understood, but for which there is no effective treatment, whereas psychiatry concerns itself with conditions for which the neural bases are unknown, but for which there do exist treatments (either talk therapy or medication) that are often effective. Happily, advances on many fronts are increasingly rendering this characterization obsolete.

Tip 16.4

This gets into game theory, another subfield from behavioral economics that has made its way into neuroeconomics.

(Motzkin, Newman, Kiehl, and Koenigs, 2012, p. 17348) (*Tip 16.3*). How this translates into MRI scanning of inmates in a portable scanner installed on the grounds of a prison is the subject of *Real-World Applications 16.1*.

There are many, many fascinating questions being addressed by social neuroscience. One, using neuropsychology to explore the nature of moral judgment, is the focus of *Web Clip 16.2*. An excellent and thorough summary of the breadth of this field can be found in the *Further Reading* by Adolphs (2010). Here, as has been the practice in previous chapters, we will focus on just a few specific topics, so as to be able to drill down to a level of detail that allows us to engage at a substantive level with its key themes and problems.

THEORY OF MIND

One concept that is central to many accounts of cognitive development, and of adult social cognition, is the theory of mind (ToM). As described by Gallagher

REAL-WORLD APPLICATIONS

16.1 The cognitive-affective-social-economic neuroscience of psychopathy

Psychopathy is a major societal problem. It is estimated to be a diagnosis of at least 25% of adult prison inmates in the United States, and is associated with disproportionately high incidences of violent crime and recidivism. Michael Koenigs did his doctoral and postdoctoral training studying the effects of brain damage on moral reasoning and economic decision making. Prompted by the similarity of some behavioral traits of psychopaths and the vmPFC patients with which he had considerable experience, he teamed up with clinical psychologist Joseph Newman to administer tests of economic decision making to incarcerated criminals with diagnoses of psychopathy. Their tasks, like many we considered in *Chapter 15*, comprised a series of offers. However these offers, like the “prisoner’s dilemma” that some readers may be familiar with, were designed to assess the extent to which people’s behavior in social situations can deviate from what a rational choice should be (*Tip 16.4*). Their findings (Koenigs, Kruepke, and Newman, 2010) indicated that psychopaths, like patients with vmPFC lesions, were disproportionately sensitive to offers that are perceived as unfair – i.e., how fairly one thinks one is being treated by another person, and/or one’s assessment of the motives of that person, influences how one reasons about money. The next step was to see if their brains showed evidence of abnormality in the vmPFC. Enter Kent Kiehl.

Kiehl (profiled in *Further Reading* by Seabrook, 2008), a professor at the University of New Mexico, has a mobile MRI scanner, built into a trailer that can be towed around the continent, and well-worked-out specifications for how the local contractor is to pour the concrete slab in the prison yard to provide the needed stability for the scanner. On the strength of their behavioral findings, Koenigs and Newman teamed up with Kiehl to set up his scanner on the grounds of a prison and enroll incarcerated psychopaths.

The first of their findings, illustrated in *Figure 16.3*, emphasizes the message also evident in *Figure 16.2*, which is that we can’t lose sight of the fact that no region of the brain does anything in isolation – understanding connections is central to understanding function. Here’s how Motzkin et al. (2012) set it up:

Both the amygdala and precuneus/PCC are densely and reciprocally connected with vmPFC ... and both areas have been associated with reduced activity in psychopathy ... Moreover, the interactions between vmPFC and these areas are thought to subserve key functions related to psychopathy. vmPFC–amygdala interactions are thought to underlie aspects of emotion regulation, aggression, and stimulus–reinforcement associations [we’ll get to this in *Chapter 17*], while vmPFC–precuneus/PCC interactions are thought to underlie aspects of self-reflective processing. (p. 17348)

For the study, they recruited prison inmates who did and did not have a diagnosis of psychopathy, and acquired diffusion tensor imaging (DTI; *Methodology Box 2.1*), T1-weighted MRI (*Methodology Box 2.1*), and resting-state fMRI (*Methodology Box 11.1*) scans and analyses. *Figure 16.3* illustrates the connectivity results, which do, indeed, implicate some of the tracts also estimated to have been compromised by Phineas Gage’s injury. Another analysis (Ly et al., 2012, not shown here) found evidence for cortical thinning in many areas, although none of them in vmPFC.

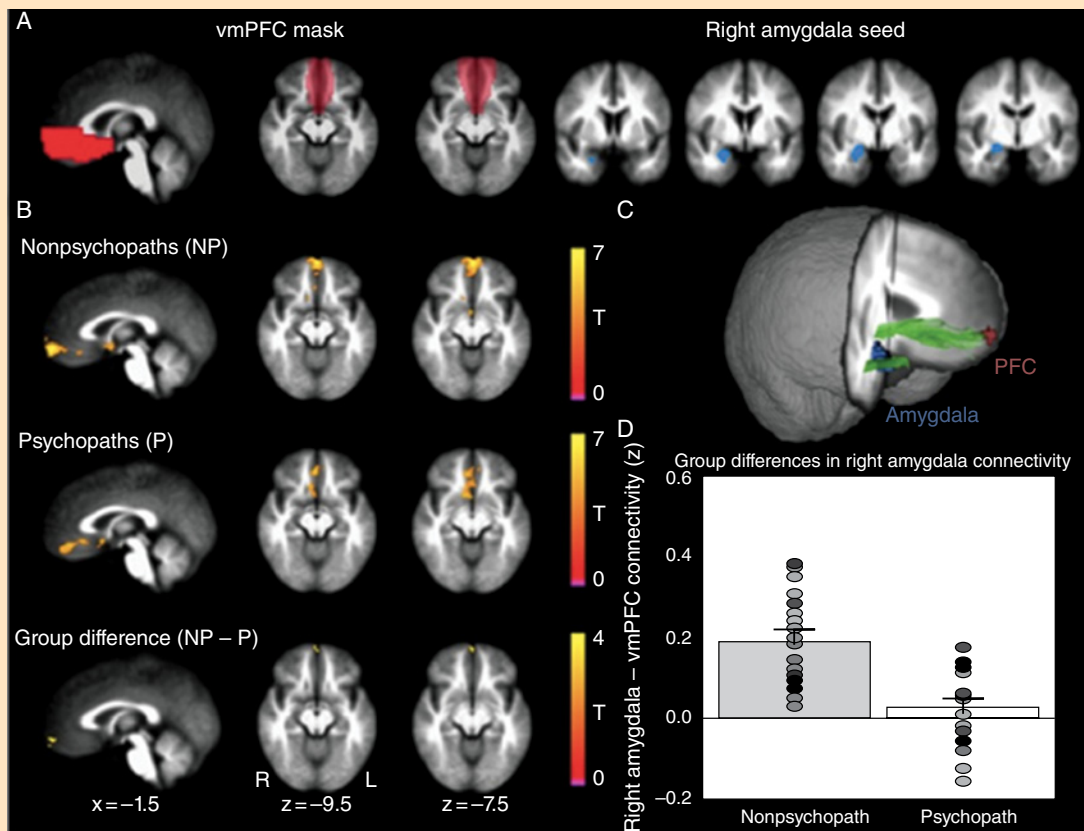
Real-World Applications 16.1 *Continued*

FIGURE 16.3 Amygdala-vmPFC connectivity differs between psychopathic and nonpsychopathic incarcerated inmates. **A** illustrates the amygdala seed and vmPFC mask used for resting-state connectivity analysis. **B** illustrates regions of significant resting-state functional connectivity in the two groups, and the difference map. Similarly decreased functional connectivity was found between medial parietal cortex and vmPFC. **C** illustrates the path of the uncinate fasciculus, the primary white-matter pathway connecting vmPFC with the anterior medial temporal lobe, shown by DTI to be compromised in the psychopath group. **D** shows individual subject values that went into B. Source: Motzkin, Newman, Kiehl, and Koenigs, 2012. Reproduced with permission of the Society of Neuroscience.

and Frith (2003), “It underpins our ability to deceive, cooperate and empathize, and to read others’ body language. It also enables us to accurately anticipate other people’s behaviour, almost as if we had read their minds” (p. 77). The ToM literature is considered in this section because the core questions underlying many claims, and debates, within this subfield of social

neuroscience have relevance to other subfields of social neuroscience and, more broadly, to other areas of cognitive neuroscience. An additional reason that this literature is of interest to many is the widely (although not universally) held view that impairments of ToM may underlie one or more variants of **autism spectrum disorder (ASD)**.

The ToM network

ToM reasoning has been theorized to require two representational steps. First, one must represent another person as an individual. Second, one must represent that person's mental states. This second stage – engaging in mentation about the thoughts, motivations, and knowledge of another – is referred to as “mentalizing.” The British cognitive neuroscientist Christopher Frith was the senior author on one of the first neuroimaging studies to explicitly study ToM mentalizing (Fletcher et al., 1995), and continues to be an influential contributor to the field. This first study, and many subsequent neuroimaging studies of ToM, used a story comprehension procedure in which subjects read brief vignettes, answering a question about each, and the contrast of interest compared activity associated with stories that required ToM mentalization vs. stories that did not. In their 2003 review, Gallagher and Frith emphasize three regions, each contributing differently to performance on theory of mind tasks. Two of these are in the temporal lobes, and the authors suggest that these aren't involved in mentalizing per se, but perhaps in functions that are necessary precursor steps. To the right posterior superior temporal sulcus they ascribe “understanding the meaning of stories and cartoons involving people, with or without the requirement to mentalize ... understanding causality and intentionality ... the attribution of intentions to the movements of geometric shapes ... and taking the self-perspective” (pp. 80–81). They note that this region has also been associated with the perception of biological motion, including movements of the hands, body, lips, and eyes. To the temporal poles they ascribe retrieval of semantic and episodic memory that can be critical for succeeding on a ToM task.

The anterior paracingulate cortex, in contrast, is argued to be recruited for tasks in which subjects take “an ‘intentional stance’, which is to treat a system as an agent, attributing to it beliefs and goals” (p. 79). This is seen, for example, when the subject is playing the game “rock, paper, scissors” (in the UK, “stone, paper, scissors”) in two conditions, one in which she believes that the opponent is a computer (in which the anterior paracingulate cortex is not engaged), the second in which she believes that the opponent is a

human being outside the scanner (in which case the anterior paracingulate cortex is engaged). (This anterior paracingulate region corresponds closely to the “medial PFC” (MPFC) region that another group routinely identifies with its “ToM localizer” task, and which will be illustrated later in *Figure 16.6*).

Two asides about anterior paracingulate cortex

First, the anterior paracingulate region implicated in ToM mentalizing by Frith and colleagues is in a similar location to the one identified by Somerville et al. (2013) when their subjects thought they were being watched while in the scanner (*Real-World Applications 15.1*).

Second, there is a fascinating variant of frontotemporal dementia (FTD), the same category of disease that produces primary progressive aphasia of the semantic subtype (a.k.a., semantic dementia), that prominently affects the medial frontal cortex, including vmPFC and anterior paracingulate cortex, in addition to a focal region of the anterior insular cortex, deep within the Sylvian fissure. It is referred to as the behavioral variant (bv)FTD, and manifests as a progressive decline in control over emotion, social behavior, personal conduct, and decision making (see *Further Reading* by Seeley, Zhou, and Kim, 2012).

Now we move on to the posterior superior temporal sulcus.

The temporoparietal junction (TPJ)

An influential, and controversial, paper from Rebecca Saxe and her then-PhD advisor Nancy Kanwisher has challenged the localization conclusions of Frith and colleagues with regard to ToM mentalizing. Saxe and colleagues have argued, instead, for a critical role for the temporoparietal junction (TPJ; the same region that Gallagher and Frith (2003) refer to as the posterior superior temporal sulcus) in ToM mentalizing. Saxe and Kanwisher (2003) noted that previous studies, including some by Frith and colleagues, relied on the practice from developmental psychology of using “false belief” stories as the gold standard for assessing ToM. An example of a false belief scenario, adapted from Gallagher and Frith (2003), is as follows: (1) *Chris knows that the chocolates have been moved from the drawer*

to the cupboard. (2) He also knows that Helen does not know that they were moved. (3) Chris knows that Helen wants a chocolate. (4) Chris predicts Helen will look for a chocolate in the _____. Children younger than age 4, as well as patients with developmental disabilities, including some patients with ASD, will mistakenly answer that Chris thinks that Helen will look in the cupboard, which is interpreted as an inability to comprehend that Helen has a false belief about the location of the chocolate. That is, an inability to understand the world from Helen's perspective. With regard to the TPJ, what Gallagher and Frith (2003) have argued, in effect, is that TPJ plays a role in proposition 3, understanding that Helen wants a chocolate, but not in understanding proposition 4.

Now back to Saxe and Kanwisher (2003). They questioned the validity of exclusive reliance on false belief scenarios to engage ToM mentalizing, and developed a procedure for an fMRI study that used many different contrasts (*Tip 16.5*). Experiment 1 contrasted ToM stories vs. mechanical inference stories and vs. human action stories. A group analysis indicated greater activity for ToM than mechanical inference stories in five regions: left and right TPJ; left and right anterior superior temporal sulcus; and precuneus. At the single subject level, activity in TPJ was greater for ToM stories vs. mechanical inference stories in 22 of 25 subjects (bilaterally in 14, left in 5, and right in 3). Importantly, for 14 subjects the authors also performed a localizer scan for the extrastriate body area (EBA; see *Figure 9.5*), and found that although this region was anatomically adjacent to the TPJ, it did not respond to ToM stories. Thus, they reasoned, TPJ does not merely respond to the physical presence of a person. Experiment 2 from Saxe and Kanwisher (2003) contrasted activity associated with (1) false belief stories, (2) false photograph stories,

(3) desires, (4) inanimate descriptions, and (5) physical people. Desire stories were designed to describe the character's goals or intentions, and thereby to also rely on ToM mentalizing. The findings from experiment 2 largely mirrored those from experiment 1, with the additional fact that desire stories also produced elevated activity in TPJ.

In a follow-up study, Saxe and Powell (2006) note in the Introduction that:

Research suggests that infants and young children understand mental-state concepts like desires, goals, perceptions, and feelings ... However, not until the age of 4 do children seem to understand concepts like belief – concepts that require understanding that mental representations of the world may differ from the way the world really is. Developmental psychologists have therefore hypothesized that the later-developing system for representing the specific (representational) contents of mental states, such as beliefs, is a distinct component of human understanding of other minds. (p. 692)

Thus, for this study, they hypothesized that if the late-developing component of ToM – what we've been calling "ToM mentalizing" – remains distinct in adulthood, then they would identify one or more brain regions recruited for ToM mentalizing stories, but not for stories invoking mental-state concepts such as those listed in the previous quote. Their findings indicated that TPJ responded specifically to ToM mentalizing stories, whereas more posterior regions, in the supramarginal gyrus of the inferior parietal lobule, responded preferentially to stories invoking mental state concepts.

Challenge to the specificity of TPJ for ToM

Why, at the beginning of this section, did I characterize this work as controversial? One reason relates to the point that has been made about the constructs of *conscientiousness*, and, earlier in this chapter, *trustworthiness* – many feel that such complex and nuanced mental constructs are very unlikely to be localizable to just one or a small handful of regions. At a psychological level, Decety and Lamm (2007) have noted that "Most if not all of the concepts used in social psychology emanate from our folk psychology; we are prisoners of words. Theory of mind, perspective-taking, empathy – these are all complex

Tip 16.5

Note that, although some of the ToM studies from Frith's group have used the false belief task, the "rock, paper, scissors" task of Gallagher, Jack, Roepstorff, and Frith (2002) certainly did not. Saxe and Kanwisher (2003) do not cite Gallagher et al. (2002).

psychological constructs that cannot be directly mapped to unique single computational mechanisms, and the functions that they attempt to describe [must be] underpinned by a network of areas" (p. 584). From a different perspective, that of systems neuroscience, the portrayal of TPJ as specific for ToM mentalizing smacks of a strongly modular view of the brain that's almost phrenological.

What's at stake in this debate isn't just whether ToM mentalizing "activates" area X or area Y in the brain, but a deeper question that relates to the age-old question of "nature vs. nurture." From the "nature" perspective, evidence that the TPJ may be specialized for ToM mentalizing is taken by some to be consistent with nativist theories in cognitive development, and the idea that ToM may be a trait with which we are born, and which matures to take its place in the mind at around 4 years of age. From the "nurture" perspective, psychological abilities like ToM mentalizing are portrayed as emerging, through learning, from a general-purpose architecture that's neither specialized for, nor specific to, any one function. This latter perspective is also articulated by Decety and Lamm (2007): "activation in the TPJ during social cognition may ... rely on a lower-level computational mechanism involved in generating, testing, and correcting internal predictions about external sensory events (p. 583)." More specifically, Decety and Lamm have proposed that "social cognition has gradually arisen from general pervasive perception-action coupling mechanisms" (p. 580) and that a function of the TPJ may be "compar[ing] internal predictions with actual external events" (p. 581).

Note that, in many ways, these objections are reminiscent of some that we considered in relation to strong claims that the fusiform face area (FFA) is specialized for face processing (*Chapter 9*). Indeed, a 2008 paper from "social cognitive and affective neuroscientist" Jason Mitchell is clearly motivated by his concerns about strong claims of specificity for activity in TPJ. In the paper's Introduction, Mitchell notes that

The preferential response of this region during belief stories has prompted claims that "temporo-parietal junction is selectively involved in reasoning about the contents of other people's minds" (Saxe and Kanwisher 2003, p. 1835); "the response of the RTPJ [right TPJ] is

highly specific to the attribution of mental states" (Saxe and Wexler 2005, p. 1397); and "the BOLD response in the RTPJ is associated with a highly specific cognitive function ... the ability to attribute thoughts to another person" (Saxe and Powell 2006, p. 697).

He then goes on to say:

Although this research has done an exhaustive job of demonstrating that RTPJ contributes to social cognition only when perceivers must infer the beliefs of another person, it has, importantly, neglected substantial evidence that this region may also subserve a set of attentional processes that are not specific to social contexts. In a literature that has developed in parallel with the work of Saxe and colleagues, researchers have repeatedly observed increased RTPJ activity when perceivers must break their current attentional set to reorient to task relevant stimuli. (p. 262)

What Mitchell is referring to here is the "circuit breaker" in the Corbetta and Shulman (2002) model that we reviewed in some detail in *Chapter 8*. In this regard, Mitchell (2008), Decety and Lamm (2007), and others have noted that right TPJ is a key region whose damage often produces hemispatial neglect (see, e.g., *Figure 6.4*).

To investigate this overlap, Mitchell (2008) closely followed procedures used previously by Saxe and colleagues, most notably using Saxe's "ToM localizer" task (which she shared with him for this purpose), and then interrogating the so-identified "ToM regions" with an attention cuing task. The ToM localizer contrasts story comprehension for false belief scenarios that require ToM mentalization with story comprehension for "false photograph" scenarios, the latter requiring subjects to represent the false content of a physical representation, such as a photograph or map, that doesn't contain any "person" content. The spatial cuing task is a speeded target detection task in which a cue that precedes the target by a variable period of time informs subjects where the target will appear. On a small proportion of trials, the cue is invalid, thereby requiring a reorientation of attention to the unexpected target location (Posner, 1980).

Mitchell's finding was that the TPJ, but not other regions identified by the ToM localizer (medial PFC, precuneus, superior frontal gyrus), showed elevated responses on all trials of the attentional cuing task, and

disproportionately large responses on invalidly cued trials. His conclusion was unflinching:

These results not only undermine the assertion that RTPJ is selective for inferring the beliefs of other people, but are also incompatible with any strong claim that this region participates selectively in social cognition, broadly construed. Greater activation in RTPJ was observed on a simple attentional cueing task in which no other people were relevant to the task: participants merely pressed 1 of 2 keys to indicate the location in which a target asterisk appeared. In so demonstrating that RTPJ activity is not selective for social cognition, the current results suggest that theory-of-mind and attentional reorienting may both require a solution to the same computational problem. That is, the observation that the same RTPJ region is engaged by these 2 tasks suggests the recruitment of a cognitive process that contributes to both belief attribution and attentional reorienting. Although the current study does not specify the exact nature of this shared process, it does suggest the need to develop a better conceptual account of what computational problems might jointly be faced by belief attribution and attentional reorienting and, subsequently, to test hypotheses about the processes that are deployed by the human mind to solve this particular cognitive challenge. (p. 268)

False beliefs (?) about Rebecca Saxe's mind

Not surprisingly, Saxe objected to this interpretation. Although this author does not have any direct knowledge about the reasoning and motivation behind all the studies that have subsequently come out of her lab, at this point it will serve a useful didactic purpose for us to engage in theorizing about Rebecca Saxe's mind – even if it means entertaining a false belief or two ;-). Specifically, it will be instructive for us to consider three studies published since 2009 in terms of how effectively they support her argument that the TPJ plays a privileged role in ToM mentalization, in the context of the Mitchell (2008) results. The first is a replication of the Mitchell (2008) study, but at a higher spatial resolution. The second is a study employing repetitive transcranial magnetic stimulation (rTMS) to the TPJ while subjects perform a moral judgment task. The third is a study applying multivariate pattern analyses (MVPA) to fMRI data

from neurotypical adults and adults with ASD performing a moral judgment task. We will consider each in turn, because they are instructive with regard to inferences about specificity that can be supported by different experimental approaches and different kinds of data.

fMRI with higher spatial resolution

This first of the three studies actually doesn't require any mentalizing on our part, because Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, and Saxe (2009) explicitly frame it as a replication of the methods of Mitchell (2008), but in a scanner featuring a higher magnetic field (3 tesla, as opposed to the 1.5 tesla scanners that Mitchell and Saxe's group had previously used). The higher magnetic field strength allowed Scholz et al. (2009) to acquire higher resolution data ($1.6 \times 1.6 \times 2.4$ mm voxels; as contrasted with the $3.75 \times 3.75 \times 6$ mm of Mitchell (2008), and of previous studies from Saxe's group). The empirical results, some of which are illustrated in *Figure 16.4*, were equivocal. First, Scholz et al. (2009) replicated Mitchell's (2008) finding that ToM-defined regions were sensitive to the valid–invalid contrast in the spatial attention task, and that spatial attention-defined regions were sensitive to the ToM contrast. In a second analysis, they determined that there was only a small amount of overlap (<10%) in voxels identified in the ToM localizer and the attention reorientation localizer. In a third analysis, they found no evidence of correlation of voxel activity across the two tasks. Finally, they evaluated whether there was reliable spatial separation between the “peak activations” for the two tasks, and determined that there was.

Let's take these findings in (roughly) reverse order. To a certain extent, the second, third, and fourth each depend on there being high test–retest reliability with fMRI. This is because interpretation of a finding that condition A and condition B do not activate precisely the same voxels during a scanning session is only meaningful if one knows that condition A at time 2 would activate the same voxels as had condition A at time 1, and that condition B at time 2 would activate the same voxels as did condition B at time 1. If, however, the fMRI signal is not highly reliable with regard to the specific voxels that are activated from session to session, then showing that condition A and condition

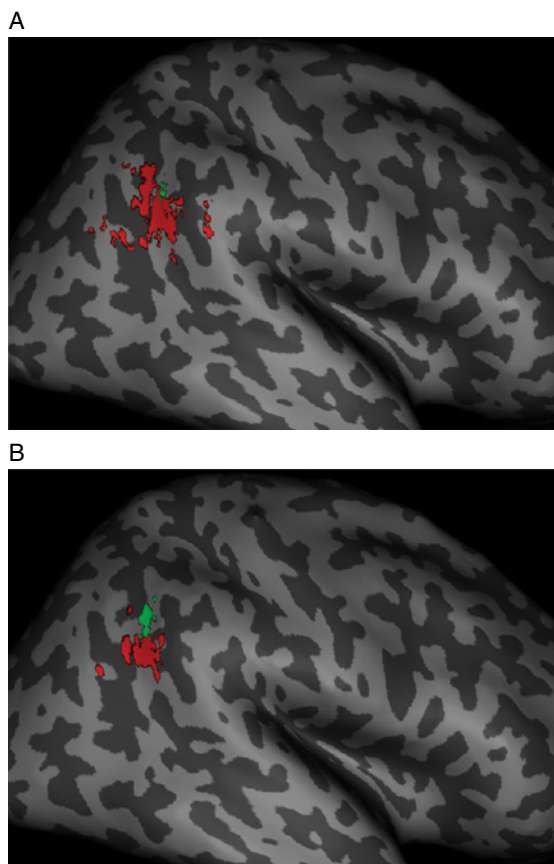


FIGURE 16.4 Activity associated with processing false belief stories (i.e., ToM mentalizing; red voxels), and with invalid spatial attention cues (i.e., attentional reorienting; green voxels). **A** shows group analysis results, with the same threshold applied to the statistical maps for each. **B** shows results from the same statistical contrasts, but with the threshold for the attentional orienting lowered, such that there would be roughly the same number of voxels for each contrast (82 for ToM mentalizing; 72 for attentional reorienting). Source: Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, and Saxe, 2009. Reproduced with permission of the authors.

B both activate the same region, but different specific voxels within that region, does not constitute strong support for the conclusion that the two conditions activate strictly nonoverlapping regions, because the nonoverlap resulting from these two scans could just be due to chance. To this author's knowledge, not

many such assessments of test–retest reliability have been performed with fMRI. In one study that did, my then-postdoctoral fellow Eva Feredoes compared activation maps from a button pushing task and a verbal working-memory task (Feredoes and Postle, 2007). She found a range from 15.2% to 6.5% of the same voxels appearing in thresholded statistical maps, depending on the particular contrast, when subjects performed the same tasks in different sessions. This suggests that the precise spatial stability of univariate fMRI maps may be relatively low, and therefore Scholz et al. (2009) would have needed to have empirically tested the overlap of $\text{ToM}_{\text{Time 1}}$ vs. $\text{ToM}_{\text{Time 2}}$ and of $\text{Spatial Attention}_{\text{Time 1}}$ vs. $\text{Spatial Attention}_{\text{Time 2}}$ for the results of their analyses to have supported a strong interpretation.

Turning to the first analysis, it replicated the findings of Mitchell (2008), and was, therefore, consistent with his conclusion that the ToM-defined region located in the vicinity of TPJ is not specific for ToM.

rTMS of TPJ during moral judgments

Moral judgment is another domain through which one can study social cognition. Koster-Hale, Saxe, Dungan, and Young (2013) have described it as such:

The distinction between intentional and accidental acts is particularly salient in the case of moral cognition. Adults typically judge the same harmful act (e.g., putting poison in a drink, failing to help someone who is hurt, making an insensitive remark) to be more morally wrong and more deserving of punishment when committed intentionally vs. accidentally ... These moral judgments depend on individuals' ability to consider another person's beliefs, intentions, and knowledge, and emerge relatively late in childhood, around age 6–7 y. (p. 5648)

In a 2010 study led by Saxe's then-postdoctoral fellow Liane Young, and including TMS pioneer Alvaro Pascual-Leone, Saxe's group used TMS applied via two different protocols to test for evidence that right TPJ plays a causal role in moral reasoning. Their TMS procedures are described in *Methodology Box 16.1*.

The study (Young, Camprodon, Hauser, Pascual-Leone, and Saxe, 2010) crossed three factors in each of two experiments: intention of the protagonist (malicious, neutral); outcome suffered by the recipient of the action (harm, no harm); TMS (right TPJ, control

METHODOLOGY BOX

16.1 Using transcranial magnetic stimulation (TMS) to alter neural functioning

The first stage in the procedure of Young et al. (2010) was to identify the brain regions that they wanted to target with TMS by scanning subjects while they performed the ToM localizer. Next, on a subsequent day, the subject came to the TMS laboratory, where a digitized 3-D representation of his head (skull and brain) was stored on a computer. After positioning the subject in the chair, he was fitted with infrared markers that allowed for the coregistration of his skull with the computer's digital image of the skull. As a result, when the experimenters touched a part of the subject's skull with an infrared-detectable wand, the computer image of the subject's skull would also be "touched" by an image of a virtual wand. Next, they could virtually "peel away" the skull to visualize where the wand was located relative to the subject's brain. Thus, with knowledge of the shape of the magnetic field produced by the "wand" (i.e., from the TMS stimulator), the researchers could position the TMS coil on the scalp such that its pulses would stimulate the area that they want to target. This approach is sometimes referred to as "frameless stereotaxy," in that it uses the same logic as stereotaxic surgery. In stereotaxic surgery, a patient is scanned while wearing a cage-like frame over her head, and the subsequent surgery is also performed while she is wearing the cage, so that the location of unseen brain regions can be calculated in reference to the invariant spatial relation between the brain and the frame.

Young et al. (2010) used two TMS protocols, which they referred to as "offline" and "online." The logic of offline TMS is to deliver a prolonged train of TMS pulses delivered at 1 Hz, in their case for 25 minutes. This stimulation protocol is known to produce a transient "virtual lesion" in the targeted region, in that cerebral metabolism, as measured by PET, will be markedly lower in the stimulated tissue for a duration of time roughly equivalent to the duration of stimulation. Although the mechanism is not fully understood, prolonged 1 Hz TMS reduces the excitability of the affected tissue, such that it will not function normally. This approach is "offline" in the sense that low-frequency TMS is delivered first, then the subject performs the task during the ensuing 25 minutes – TMS is not delivered simultaneously with task performance.

The logic of "online" TMS, in contrast, is analogous to that of microstimulation protocols that we have considered previously: it is delivered simultaneous with task performance and, importantly, can target specific epochs of a trial, leaving the others unaffected. Although there is no hard-and-fast rule for naming conventions, TMS delivered at a frequency higher than 1 Hz is generally referred to as repetitive (r)TMS. Unlike prolonged 1 Hz TMS, the effects of rTMS varies, sometimes disrupting behavior, sometimes facilitating it. The factors that determine this are still being worked out. Importantly, however, so long as there is adequate spacing between trains of rTMS (roughly 10 seconds or so between trains), there is not a build-up of effects on brain activity (Hamidi, Johnson, Feredoes, and Postle, 2011).

area). In both experiments, the story that each subject heard included four sections – "background," "fore-shadow," "intention," "action" – followed by a judgment in which the experimental subject rated the protagonist's actions on a permissible-to-forbidden rating scale. The inclusion of an "active stimulation"

control region, which for this study was a region of parietal cortex located 5 cm posterior to the TPJ, allowed the authors to evaluate whether any result that they might find by applying rTMS to TPJ could be explained as a nonspecific result of simply applying magnetic stimulation to the brain. (Another type of

control that is sometimes used in TMS studies, sham stimulation that produces the sounds and tapping sensation of TMS, is not as effective a control, because it does not deliver any stimulation to the brain.)

Both studies (offline and online; *Methodology Box 16.1*) found the same result, which is that TMS delivered to the TPJ selectively altered judgments on the malicious-intent stories for which the outcome was no harm. For these stories, rTMS to the TPJ, but not the control area, had the effect of making subjects judge the action of the protagonist as being more “permissible” than they did when TMS was delivered to the control area. In the online study, a 500 ms train of 10 Hz TMS was delivered simultaneous to the onset of the moral judgment question, meaning that this effect could be isolated to the judgment itself, not to some other factor, like comprehension of the protagonist’s intention.

These findings, therefore, revealed a necessary contribution of TPJ and, possibly, other regions functionally coupled with it (i.e., having elevated functional connectivity with TPJ), for moral reasoning. It is particularly striking that the TMS effects were seen in just this one cell of the experimental design, meaning, among other things, that they did not also affect judgments for scenarios when the protagonist had ill intent *and* something bad happened. This suggests that TPJ contributes most importantly to assessing morally ambiguous situations, in which there is a mismatch between intent and outcome.

Note, however, that the results of Young et al. (2010) do not speak to the possible specificity of this region’s function in relation to moral judgment versus other aspects of cognition. In particular, the authors did not assess the effects of rTMS on attentional reorienting. Indeed, there have been several studies of spatial attention (that did not test ToM) that have shown clear effects on attention when TMS is delivered to TPJ. Thus, it’s a reasonable assumption that, had Young et al. (2010) included a test of attentional reorienting, they would have found effects on this task, too. (This observation, in turn, must be qualified by the fact that the volume of tissue affected by TMS is sufficiently large – on the order of a cubic centimeter or so – that if the true organization of the brain were such that ToM functions and attentional reorienting functions were segregated in a strict but anatomically

adjacent way (as portrayed, e.g., in *Figure 16.4*), the coarse resolution of rTMS would make it impossible to target just one of these regions.)

Moral judgments, TPJ, and autism spectrum disorder

The third of the three post-Mitchell (2008) studies that we will consider here is a study scanning ASD individuals while performing a moral judgment task, and comparing their performance and brain activity to those of neurotypical (NT) individuals (*Tip 16.6*). Its results are detailed in *Research Spotlight 16.1*. Finally, after so cavalierly theorizing about Rebecca Saxe’s mind, it’s only fair that we let her speak for herself, as she does at *Web Link 16.2*.

A final assessment of the role of TPJ in ToM mentalization

The TMS study of Young et al. (2010) and the fMRI study of Koster-Hale et al. (2013) each provide strong evidence that right TPJ is involved in making moral judgments. This is not equivalent, however, to saying that they have shown it to be a “moral judgment area.” Indeed, the findings of Mitchell (2008) and of Scholz et al. (2009) both suggest that voxels of TPJ recruited for ToM mentalization are also recruited for the reorienting of spatial attention. The findings from these two individual studies are consistent with the results of a meta-analysis of neuroimaging studies by Decety and Lamm (2007). The term “meta-analysis” refers to an analysis of the results from many other studies. It’s “meta” because one isn’t collecting data from individual subjects and analyzing these data, but, instead,

Tip 16.6

The decades-old convention in neuropsychology is to refer to control subjects as “normal controls” or “healthy controls.” In the context of developmental disorders, however, particularly ASD, such labels are considered pejorative, because they imply that the patients are not “normal” or “healthy.” Thus, the more neutral “neurotypical” (NT) is the label commonly used in such studies.

RESEARCH SPOTLIGHT**16.1 ToM-related BOLD signal intensity and MVPA, in neurotypical and ASD adults**

With regard to moral reasoning, the behavioral profile of high-functioning individuals with ASD has been summarized as follows by Koster-Hale et al. (2013):

Individuals with ASD do not have impairments in moral judgment as a whole: children with ASD make typical distinctions between moral and conventional transgressions ... and between good and bad actions ... However, they are delayed in using information about innocent intentions to forgive accidents ... Furthermore, even very high-functioning adults with ASD who pass traditional tests of understanding (false) beliefs neglect beliefs and intentions in their moral judgments compared with NT adults. (p. 5651)

Koster-Hale et al. (2013) assessed reasoning about beliefs and intentions with stories similar to those used in the TMS study of Young et al. (2010), with the exception that information in this study was provided in the order of “background,” “action,” “outcome,” “intent.” High-functioning individuals with ASD were recruited via advertisements placed with a regional Asperger’s Association. They were rated as scoring higher on the Autism Quotient questionnaire than NT subjects, and received a diagnosis of ASD from a clinician.

Behaviorally, subjects in both the ASD and NT groups judged intentional harms to be more blameworthy than accidental harms, and judged both of these to be worse than neutral acts. The two groups differed, however, in that ASD subjects assigned more blame for accidental harms than did NT adults, and they assigned less blame for intentional actions.

In both groups, univariate analyses of fMRI signal intensity, averaged over the whole trial, revealed higher responses for harmful than neutral actions in all four ROIs illustrated in *Figure 16.5*, panel A. However, the two groups differed when just the final 8 seconds of the trial were considered. This was the point in the trial when the intention of the protagonist was revealed to be malicious or neutral. During this portion of the trial, the response in the right TPJ of NT adults was higher for accidental than intentional harms, but there was no such difference in the ASD group. Findings from MVPA analyses largely paralleled those from the univariate analyses, indicating that for both groups and in all ROIs, patterns of activity discriminated between harmful vs. neutral acts. Also mirroring the univariate analyses, it was only in NT subjects that patterns of activity in right TPJ discriminated reliably between accidental and intentional harms. Further analyses, illustrated in *Figure 16.5*, panel C, indicated that for these NT subjects the strength of the multivariate discriminability of intentional vs. accidental harms in right TPJ activity predicted how strong was the behavioral distinction that that subject would make between these types of harms. That is, individuals whose right TPJ evinced more distinct patterns between intentional and accidental harms also drew a larger distinction between intentional and accidental harms in their moral judgments.

Although one might be tempted to see parallels between the TMS study of Young et al. (2010) and the fMRI findings of Koster-Hale et al. (2013), they differ importantly in that TMS of TPJ had the effect of making subjects judge failed attempts to do harm more leniently. In the present study, however, MVPA of TPJ did not distinguish between neutral acts and malicious acts that failed to harm.

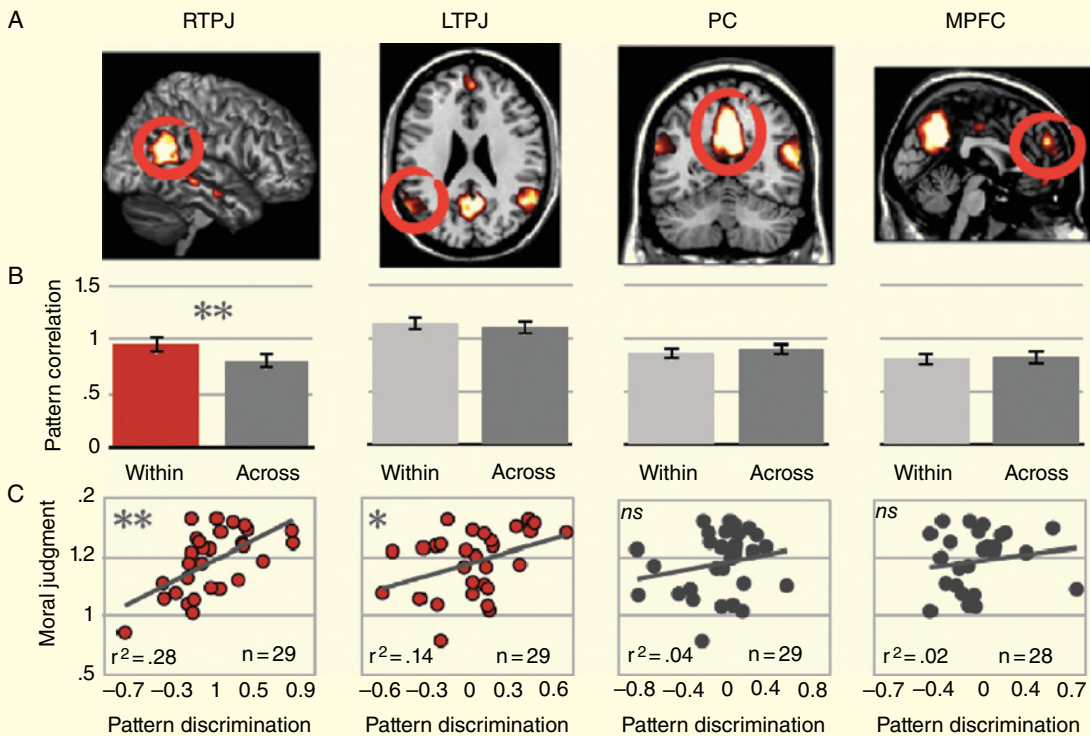
Research Spotlight 16.1 *Continued*

FIGURE 16.5 fMRI of judgment of accidental and deliberate harms. **A** shows the four regions that are routinely identified with Saxe's ToM localizer task. **B** illustrates, for each ROI, MVPA results. These analyses were performed as the correlation of multivariate patterns within each condition (i.e., correlation of the pattern from half of the *accidental harm* trials vs. the other half of the *accidental harm* trials, and of half of the *intentional harm* trials vs. the other half of the *intentional harm* trials) vs. across the two conditions (i.e., correlation of *accidental harm* trials vs. *intentional harm* trials). **C** illustrates the relation between the magnitude of the *Within* vs. *Across* difference from the MVPA ("Pattern discrimination") and how differently that person assigned blame on accidental vs. intentional harm trials ("Moral judgment"). Source: Koster-Hale, Saxe, Dungan, and Young, 2013. Reproduced with permission of PNAS Group.

assembling the results of many individual experiments, and, in effect, treating each experiment as a data point in an analysis of analyses. For neuroimaging data, the goal is often to obtain a measurement that is less susceptible to sampling error in the precise localization of effects that would be characteristic of any one fMRI study. (I.e., the problem of low test-retest reliability that was raised earlier in this section.) The

meta-analysis of Decety and Lamm (2007) included 18 studies of spatial reorienting, 15 studies of agency (the feeling of being the cause of one's own thoughts and actions), 13 studies of empathy, and 24 studies of ToM. The finding was that, in the region of the right TPJ, results from the four kinds of studies largely overlapped. Thus, it seems to be that although the right TPJ contributes importantly to moral judgment

Tip 16.7

Note that from quoted statements from both the Decety and Lamm (2007) and Mitchell (2008), one might get the impression that TPJ “can only do one thing,” and that their findings require that this be a process that is involved in both ToM and attentional reorienting. However, in higher-level cortex, individual neurons and networks can contribute to many different functions. Additionally, because individual voxels contain hundreds of thousands of neurons, it almost has to be the case that the tissue within any given voxel supports a multiplicity of functions (Geoffrey Aguirre Lab, n.d.).

and perhaps also to ToM mentalization, it is not specialized for these functions. Therefore, the question about how these aspects of social behavior are produced by the brain remains uncertain (*Tip 16.7*).

OBSERVATIONAL LEARNING

In the final section of this chapter we’ll consider another aspect of social behavior, which is learning by observing the behavior of others. Although we learn many things about how the world works through direct trial-and-error, it can often be more efficient, if not safer, to watch how others approach a novel situation and, as the expression has it, “learn from their mistakes.” Formally, these two types of learning are called “experiential” and “observational,” and the latter is inherently social. On an intuitive level, it makes sense that we can learn from observing the experiences of others, and each of us can immediately think of several examples. (Here’s a simple one. You are hungry and go to the lobby in your building where the vending machines are located. There are two machines, and as you arrive you see that someone else has put money into machine 1, and is now cursing and kicking the machine because it took her money but didn’t deliver the item that she selected, nor will it return her money. As a result of this observation of the behavior of another, which machine are you going to use, 1 or 2?) As we shall

now see, however, the way that observational learning works can be nonintuitive.

In this chapter’s final study that we will consider in detail, Burke, Tobler, Baddeley, and Schultz (2010) scanned an individual while s/he performed a probabilistic learning task. On each block of trials, three pairs of randomly selected fractal stimuli were presented, and the subject had to learn, for each pair, which of the two was associated with a positive outcome, and which with a negative outcome. (On “gain” blocks, the positive outcome was gaining 10 points for selecting the correct stimulus; 0 points for the incorrect one; on “loss” blocks the positive outcome was losing 0 points for selecting the correct stimulus, and losing 10 points for selecting the incorrect one.) To capture some of the “noise” and variation that exists in the natural world, these reward contingencies were probabilistic, rather than absolute: during *gain* blocks, the “good” stimulus delivered the reward on 80% of trials, but the contingencies were reversed on the remaining 20% of trials; during *loss* blocks, the same probabilities applied. Trial types were blocked, such that the subject performed a series of *gain* trials during one scan, then a series of *loss* trials during the next.

Here’s the social/observational part. On the day of the scan, it was arranged that a few minutes after the true research subject arrived to be scanned, a second individual who was actually a part of the research team (a “confederate”) arrived and also sat in the waiting room, pretending to be a second research subject. A second experimenter then came into the room and explained to both that they would each be playing the same casino-style gambling-machine game – trying to win the most points and lose the least. It was also explained that while playing the game, the true subject would be in the MRI scanner and the confederate in a testing room, and that each would be able to observe how the other was playing. Then a digital photo of each was taken. The two players would take turns, and the face of either player would be displayed on the screen while it was that person’s turn to play. The sequence of a trial is illustrated in Figure 16.6.A–C. In addition to the blocking of *gain* and *loss* trials, the critical experimental factor was the amount of social information available to the research subject: none (*Individual learning*); some (*Action only*, in which the

subject could observe the confederate's choice, but not the outcome of that choice); and all (*Action + outcome*). One might expect (at least I did) that observing only another's choice, but not the outcome of that choice, would be of little-to-no use to an individual. How could one "learn from another's mistakes" without the knowledge of whether or not that person's choice did, or didn't, turn out to be a mistake or a

correct choice? (In the vending machine example, if you watched the person put her money in the machine, but then stepped outside to make a phone call and, when you returned to the lobby, she was no longer there, how could having seen her put her money into the machine, but not having seen her lose it, possibly help you make a better decision?) For the moment, we'll simply demonstrate that one can learn

FIGURE 16.6 Experiential and observational learning. Source: Burke, Tobler, Baddeley, and Schultz, 2010. Reproduced with permission of PNAS Group.

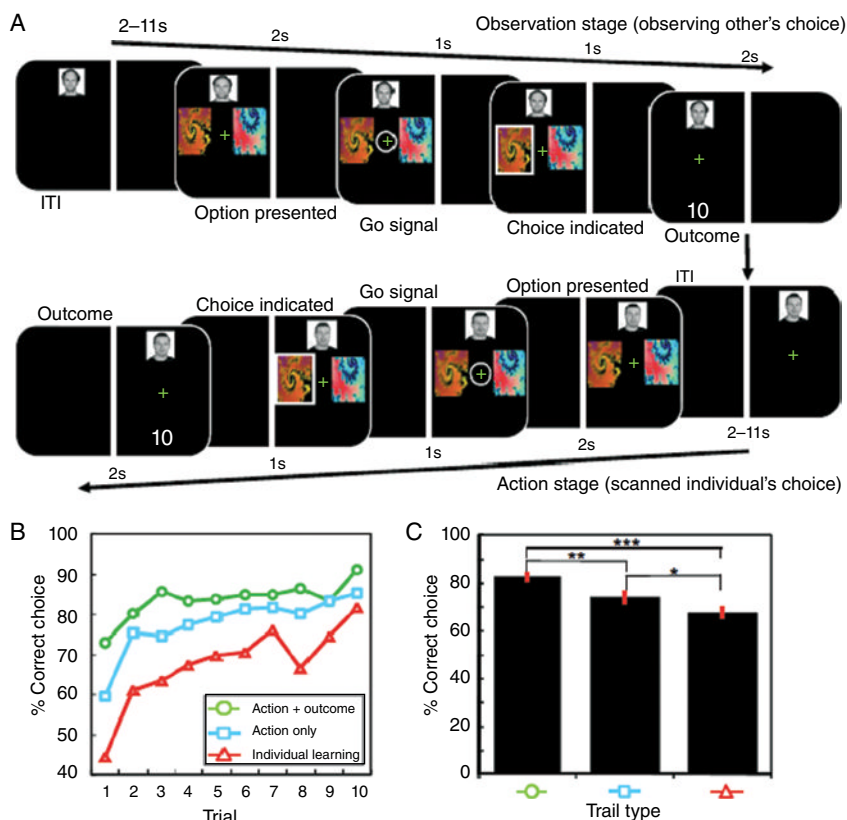


FIGURE 16.6. A–C Panel **A** illustrates the series of events on a single trial. First, in the *Observation stage*, the subject watched as the confederate's options were presented, then as the confederate received the go signal, made his choice, and was informed of the outcome (in this illustration he won 10 points). Next, in the *Action stage*, the subject went through the same sequence. On *Individual learning* trials, the subject only experienced the *Action stage*; on *Action only* trials, the subject observed all the sequences of the confederate's trial (i.e., the *Observation stage*), with the exception that the outcome was not presented. During each session, the same pair of stimuli was presented for each *Individual learning* trial, a different pair for each *Action only* trial, and still a different pair for each *Action + outcome* trial. Panel **B** illustrates, with a single subject's performance, that even beginning with the first trial, having knowledge of another's action, even without knowing what the outcome of the action was, aided performance markedly. Panel **C** shows aggregated group performance in the three learning conditions.

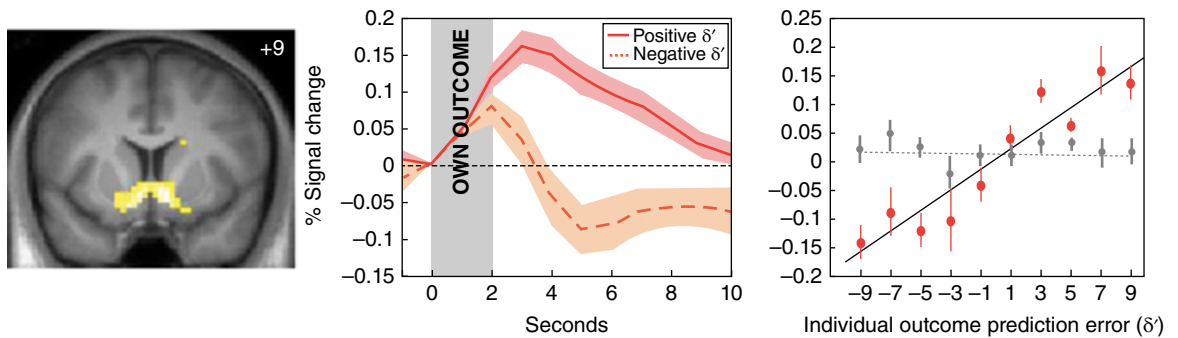


FIGURE 16.6.D The region correlating with reward prediction errors from the model was the nucleus accumbens of the ventral striatum. The center panel shows activity from this region, time-locked to the *Outcome* epoch of the *Action stage* from panel A, sorted into trials estimated by the model to produce a positive reward prediction error or a negative reward prediction error. (Recall, from *Figure 14.4*, that an unexpected reward yields a positive prediction error and the absence of an expected reward yields a negative prediction error.) The panel on the right illustrates the correlation between the model's estimates of individual outcome prediction error on each of the 10 trials of the session and the fMRI response from corresponding trials, in the nucleus accumbens (red symbols). Note that, because all trial types had an *Outcome* stage, this analysis included trials from all three trial types. To demonstrate that social and nonsocial learning are integrated in this outcome prediction error in the nucleus accumbens, the gray triangular symbols, show the same regression with expected outcome prediction errors, but when social information was removed from the model on *Action only* and *Action + outcome* trials.

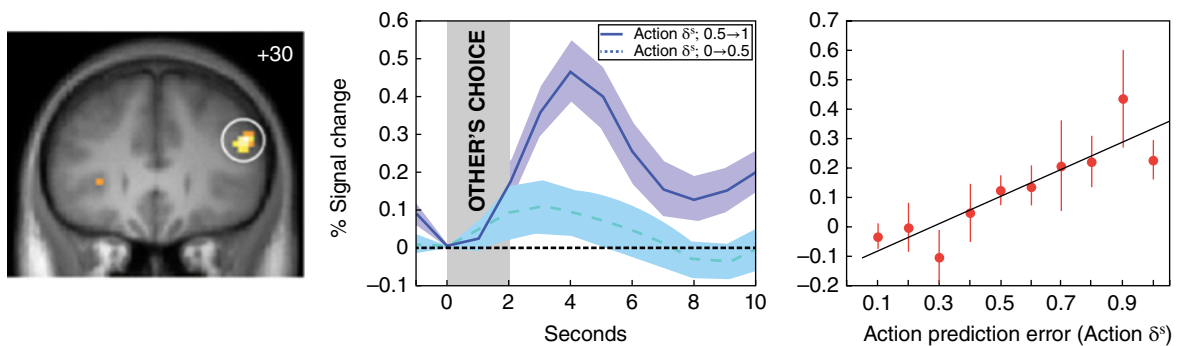


FIGURE 16.6.E The region correlating with the action prediction error signal from the model was the dorsolateral PFC. The center panel shows activity from this region, time-locked to the *Choice indicated* epoch of the *Observation stage* from panel A, sorted into trials estimated by the model to produce a large action prediction error and a small reward prediction error. The panel on the right illustrates the correlation between the model's estimate of action prediction errors and the fMRI response from corresponding trials in the dorsolateral PFC.

in these situations, then, in *Methodology Box 16.2*, we'll consider a possible explanation.

Figure 16.6.A–C shows very clearly that the *Action only* condition, in which one can observe how another behaved, but not the consequence of this behavior, produces markedly superior performance to the *Individual learning* condition. One explanation for how this may happen relies on computational reinforcement

learning algorithms. Critically, two of the inputs fed to the algorithm during each trial are new kinds of prediction error signal (*Thought Box 14.1*) whose existence was postulated by Burke et al. (2010) to account for this puzzle about learning from *Action only* trials. One of these hypothesized prediction error signals is associated with one's expectations of another's actions, regardless of what the consequence(s) of that

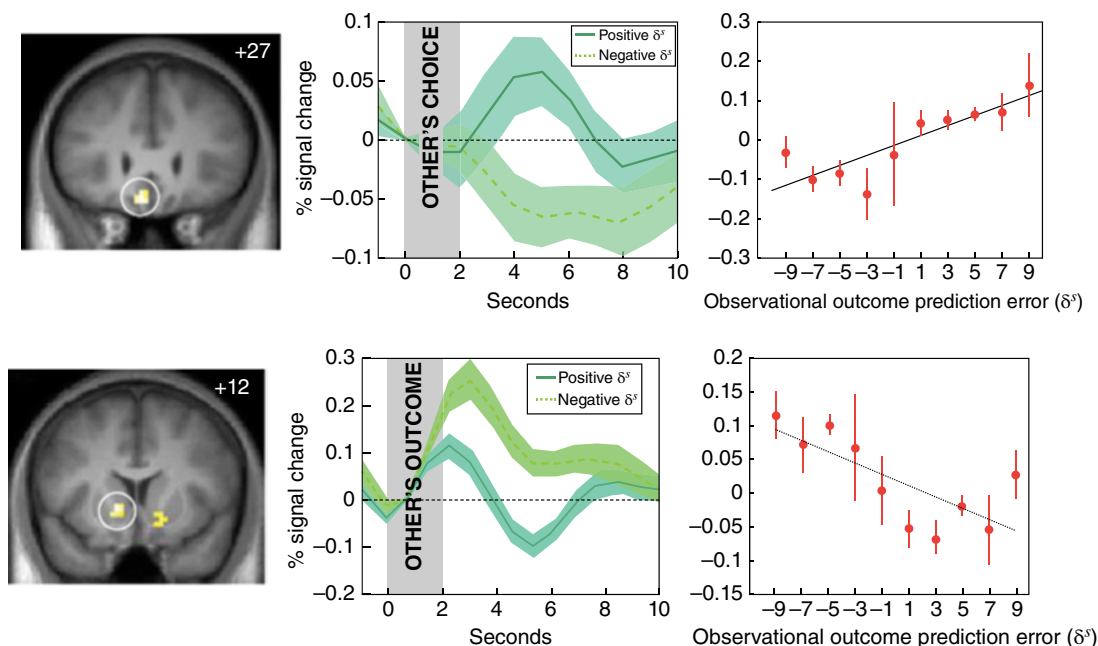


FIGURE 16.6.F The regions correlating with observational outcome prediction errors from the model were the ventromedial PFC (top row) and the nucleus accumbens (bottom row), the two responding in opposite directions. The center panel shows activity from the two regions, time-locked to the *Outcome* epoch of the *Observation* stage of from panel A, sorted into trials estimated by the model to produce a positive observational prediction error and a negative observational prediction error. The panel on the right illustrates the correlation between the model's estimate of observational outcome prediction errors and the fMRI response from corresponding trials in the two regions.

action may be. The second is associated with one's expectations about the reward that another individual will receive (thus, known as a “vicarious” reward). The computational implementation of the hypothesized mechanism is the basis for an approach to fMRI data analysis that is known as “model-based” (or, alternately, “computational”) fMRI. The rationale behind model-based fMRI is that if there's no clearly intuitive understanding of a certain pattern of behavior, conventional analyses of fMRI scans of this poorly understood behavior wouldn't provide any new insight about it. All they would show is where in the brain there are changes in activity associated with something that we don't understand. If, on the other hand, this behavioral phenomenon can be understood via a formal model, with parameters that we can understand, determining whether there are areas of the brain whose activity covaries with the parameter estimates from the model can tell us about the likelihood that the brain actually implements the computations that

are implemented by that model. (A model, after all, is just that — a *model*. It's someone's “best guess” about how a system may work.) *Methodology Box 16.2* explains the implementation of a model-based analysis that can be understood as a test of the hypothesis that three types of error prediction signal explain how learning of probabilistic reward associations differs as a function of whether that learning is purely experiential, whether it includes observed actions of another, but not the outcomes of those actions, or whether it includes observed actions and the outcomes of those observed actions.

The findings, illustrated in *Figure 16.6.D–G*, suggest that the reward prediction error signal associated with the feedback received for the subject's own response is computed in the ventral striatum. They suggest that the action prediction error signal, associated with the subject's expectation of what the confederate will do, is processed in the dorsolateral PFC. Finally, they suggest that the observational outcome prediction error signal

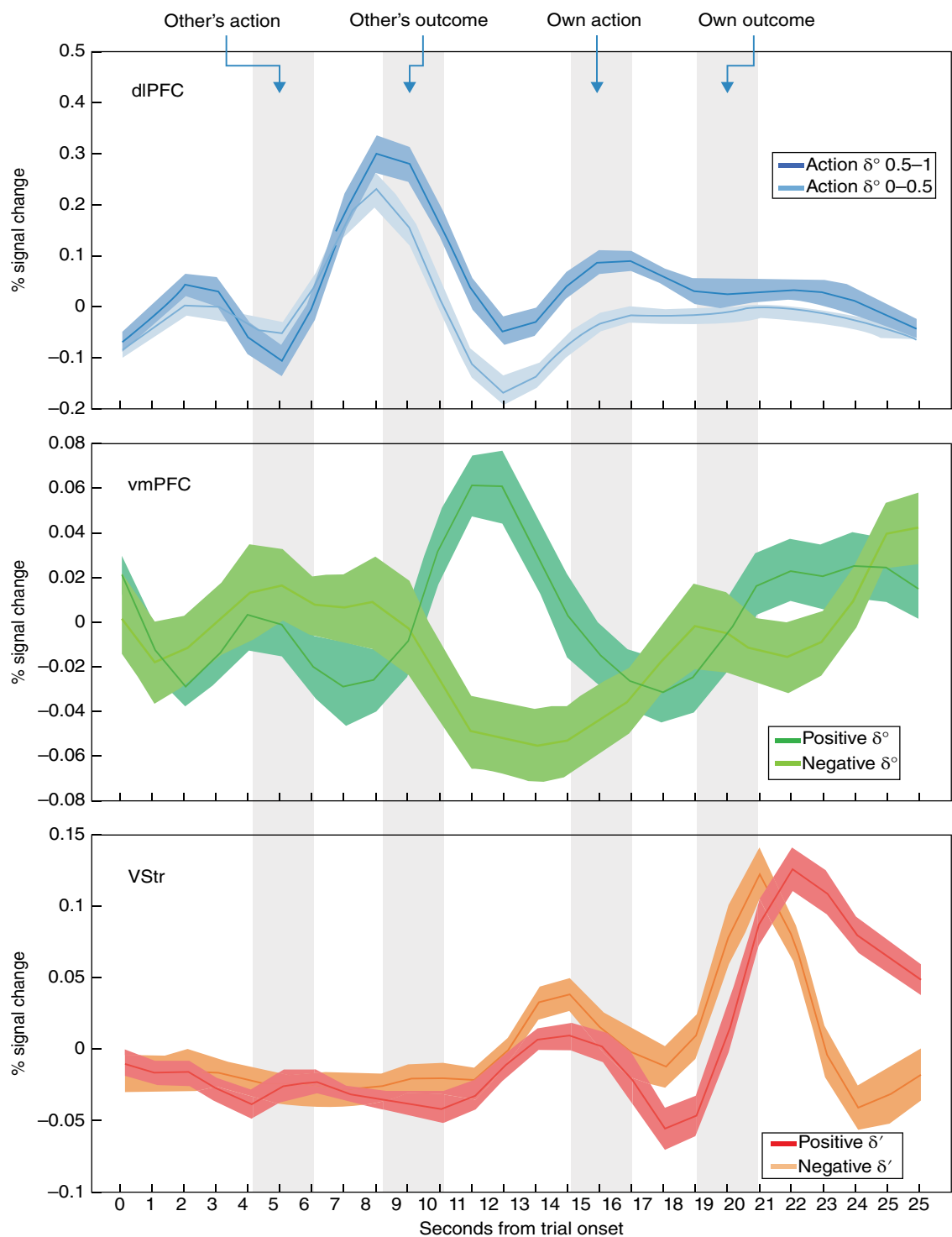


FIGURE 16.6.G The full-trial group- and trial-averaged fMRI time series from the three ROIs illustrated earlier in this figure. This panel provides a simultaneous presentation of the timing of each of the error prediction-related effects in the context of the full length of the trial, and of each of the ROIs implicated in these different facets of observation learning. ("VStr" = ventral striatum; the gross-anatomical location of the nucleus accumbens.)

METHODOLOGY BOX

16.2 Model-based fMRI

Although there is no obvious, intuitive reason why a subject should learn more quickly when he can observe the choices of another, but not the outcomes of those choices, there are computational learning algorithms that can explain it. The senior author of the Burke et al. (2010) study is Wolfram Schultz, whose discovery and characterization of the reward prediction error signal was detailed in *Thought Box 14.1*. Indeed, Burke et al. (2010) propose that human observational learning may be achieved by applying two newly hypothesized variants of prediction error signals. They implemented this idea in a computational model of probabilistic reward-learning, the “Q” learning algorithm. The mathematical details of this model are beyond the scope of this book, but what follows here is a conceptual explanation.

Initially, the authors noted that the behavior of subjects differed as a function of whether or not the outcome of the confederate’s choice was visible: on *Action only* trials (see *Figure 16.6.A–C*), they much more frequently imitated the confederate’s behavior than they did on *Action + outcome* trials. This confirmed that subjects “adaptively used available information in the different conditions,” and led the authors to reason that a process of observing another’s actions and combining this knowledge with one’s own experience “can be thought of as a two-stage learning process; first, observing the action of another player biases the observer to imitate that action (at least on early trials), and second, the outcome of the observer’s own action refines the values associated with each stimulus” (p. 14432). On *Action + outcome* trials, the authors posited an analogous two-stage process, but with the first stage corresponding to observing the outcome of the confederate’s choice.

The mechanisms that they proposed to implement the first stage of learning in these two two-stage processes were previously undescribed variants of the reward prediction error. On *Action only* trials they hypothesized that an “action prediction error” would signal the extent to which the confederate’s choice differed from what the observer expected the confederate to do. Because, unlike rewarded choices, simple motor actions do not have inherently rewarding or punishing values, they posited that action prediction errors could vary in magnitude, but not in sign. On *Action + outcome*, the hypothesized “observational outcome prediction error” would have the same properties as the reward prediction error that we have already learned about, but would be based on “vicarious” rewards, in that they pertained to the confederate, and could never have been obtained by the observer. (This is related to the concept of “fictive learning.”)

These newly proposed prediction errors can explain the observed behavior in the three trial types, but how do we know that they really exist? After all, they were “just made up” to account for this particular form of social learning. Burke et al. (2010) assessed their neurobiological plausibility by regressing the parameter estimates generated by their model, corresponding to *reward prediction errors*, *action prediction errors*, and *observed-outcome prediction errors*, against the fMRI data, to determine if the activity of any areas of the brain fluctuated on a trial-by-trial basis with the values these three parameters. The results are illustrated in *Figure 16.6.D*, *Figure 16.6.E*, and *Figure 16.6.F*.

is processed in two regions: the processing in vmPFC seems to reflect the degree to which the actual outcome of the other player was unpredicted relative to the individual's prediction, whereas the processing in ventral striatum seems to reflect the inverse of the signal. This last result means that, in ventral striatum, a positive observational outcome prediction error occurs when viewing another person lose. Although the Burke et al. (2010) model did not have an a priori prediction about this finding, the authors speculate that it may mean that "it [is] ... rewarding to view the misfortunes of others" (p. 14435).

To conclude this section, although the findings of Burke et al. (2010) can't prove, in an absolute sense, that their account is the true way that observational learning is implemented in the brain, they provide very compelling evidence that is consistent with this model.

TRUSTWORTHINESS, REVISITED

We have just reviewed the fact that neural systems in the vmPFC and paralimbic regions contribute importantly to the control of social comportment. We have also considered research aimed at understanding the neural underpinnings of ToM. Finally, we have reviewed neural evidence for an explicit computational model of social learning. Having done so, we now have a better sense of how one might go about incorporating a representation of trustworthiness into neural models of intertemporal choice. We have a better sense of what are important questions to ask, of what caveats and complications may arise. It won't be easy, or quick. It's unlikely to be uncontroversial. But now, at the end of this chapter, we have a better sense of how one might go about incorporating into our cognitive neuroscience models the very real influences of social cognition on brain and behavior.

END-OF-CHAPTER QUESTIONS

1. What are potential problems, or, at least, complications, with trying to study the neural bases of a social construct like *trustworthiness*? Why is it nonetheless important to confront these problems/complications?
2. In what ways do findings from contemporary behavioral neurology support the conclusions drawn from the case of Phineas Gage?
3. Does the vmPFC act alone in the control of social behavior? What evidence supports your answer?
4. What brain regions have been implicated as part of a "ToM network"? What function(s) have been ascribed to different components of this network?
5. What are two differing accounts of the function(s) supported by the right TJP in relation to social behavior? What fundamental theoretical perspectives underlie these two accounts?
6. What are three different experimental protocols for using TMS to study neural function? (Note: only two are considered in this chapter; a third was introduced way back in *Research Spotlight 3.2.*)
7. What is a circumstance when a model-based analysis of fMRI may be preferred over the univariate or multivariate approaches that we have emphasized up until this point in the book?
8. How might it be that learning is improved when one can observe the actions of others, but not the outcomes of these actions? Be specific about the proposed neurobiological mechanism(s).

REFERENCES

- Burke, Christopher J., Philippe N. Tobler, Michelle Baddeley, and Wolfram Schultz. 2010. "Neural Mechanisms of Observational Learning." *Proceedings of the National Academy of Sciences of the United States of America* 107 (32): 14431–14436. doi: 10.1073/pnas.1003111107.
- Corbetta, Maurizio, and Gordon L. Shulman. 2002. "Control of Goal-Directed and Stimulus-Driven Attention in the Brain." *Nature Reviews Neuroscience* 3 (3): 201–215. doi: 10.1038/nrn755.
- Decety, Jean, and Claus Lamm. 2007. "The Role of the Right Temporoparietal Junction in Social Interaction:

- How Low-Level Computational Processes Contribute to Meta-Cognition." *Neuroscientist* 13 (6): 580–593. doi: 10.1177/1073858407304654.
- Feredoes, Eva, and Bradley R. Postle. 2007. "Localization of Load Sensitivity of Working Memory Storage: Quantitatively and Qualitatively Discrepant Results Yielded by Single-Subject and Group-Averaged Approaches to fMRI Group Analysis." *NeuroImage* 35 (2): 881–903. doi: 10.1016/j.neuroimage.2006.12.029.
- Fletcher, Paul C., Francesca G. Happé, Uta Frith, S. C. Baker, Ray J. Dolan, Richard S. J. Frackowiak, and Christopher D. Frith. 1995. "Other Minds in the Brain: A Functional Imaging Study of 'Theory of Mind' in Story Comprehension." *Cognition* 57 (2): 109–128. doi: 10.1016/0010-0277(95)00692-R.
- Gallagher, Helen L., and Christopher D. Frith. 2003. "Functional Imaging of 'Theory of Mind'." *Trends in Cognitive Sciences* 7 (2): 77–83. doi: 10.1016/S1364-6613(02)00025-6.
- Gallagher, Helen L., Anthony I. Jack, Andreas Roepstorff, and Christopher D. Frith. 2002. "Imaging the Intentional Stance in a Competitive Game." *NeuroImage* 16 (3): 814–821. doi: 10.1006/nimg.2002.1117.
- Geoffrey Aguirre Lab. N.d. "Number of Neurons in a Voxel." https://cfm.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel.
- Hamidi, Massihullah, Jeffrey S. Johnson, Eva Feredoes, and Bradley R. Postle. 2011. "Does High-Frequency Repetitive Transcranial Magnetic Stimulation Produce Residual and/or Cumulative Effects within an Experimental Session?" *Brain Topography* 23 (4): 355–367. doi: 10.1007/s10548-010-0153-y.
- Harlow, John Martyn. 1868. "Recovery from the Passage of an Iron Bar through the Head." *Bulletin of the Massachusetts Medical Society* 2: 327–347.
- Koenigs, Michael, Michael Kruepke, and Joseph P. Newman. 2010. "Economic Decision-Making in Psychopathy: A Comparison with Ventromedial Prefrontal Lesion Patients." *Neuropsychologia* 48 (7): 2198–2204. doi: 10.1016/j.neuropsychologia.2010.04.012.
- Koster-Hale, Jorie, Rebecca Saxe, James Dungan, and Liane L. Young. 2013. "Decoding Moral Judgments from Neural Representations of Intentions." *Proceedings of the National Academy of Sciences of the United States of America* 110 (14): 5648–5653. doi: 10.1073/pnas.1207992110.
- Lhermitte, François, Bernard Pillon, and M. Serdaru. 1986. "Human Autonomy and the Frontal Lobes. Part I: Imitation and Utilization Behavior: A Neuropsychological Study of 75 Patients." *Annals of Neurology* 19 (4): 326–334. doi: 10.1002/ana.410190404.
- Ly, M., J. C. Motzkin, C. L. Philippi, G. R. Kirk, J. P. Newman, K. A. Kiehl, and M. Koenigs. 2012. "Cortical thinning in Psychopathy." *American Journal of Psychiatry* 169 (7): 743–749. doi: 10.1176/appi.ajp.2012.11111627.
- Macmillan, Malcolm. 2000. *An Odd Kind of Fame: Stories of Phineas Gage*. Cambridge, MA: MIT Press.
- Mesulam, M.-Marsel. 1985. "Patterns in Behavioral Neuroanatomy: Association Areas, the Limbic System, and Hemispheric Specialization." In *Principles of Behavioral Neurology*, edited by M.-Marsel Mesulam, vol. 26 of *Contemporary Neurology Series*, 1–70. Philadelphia: Davis.
- Mesulam, M.-Marsel. 1986. "Frontal Cortex and Behavior." *Annals of Neurology* 19 (4): 320–325. doi: 10.1002/ana.410190403.
- Michaelson, Laura, Alejandro de la Vega, Christopher H. Chatham, and Yuko Munakata. 2013. "Delaying Gratification Depends on Social Trust." *Frontiers in Psychology* 4: 355. doi: 10.3389/fpsyg.2013.00355.
- Mitchell, Jason P. 2008. "Activity in Right Temporo-Parietal Junction Is Not Selective for Theory-of-Mind." *Cerebral Cortex* 18 (2): 262–271. doi: 10.1093/cercor/bhm051.
- Motzkin, Julian C., Joseph P. Newman, Kent A. Kiehl, and Michael Koenigs. 2012. "Reduced Prefrontal Connectivity in Psychopathy." *Journal of Neuroscience* 31 (48): 17348–17357.
- Posner, Michael I. 1980. "Orienting of Attention." *Quarterly Journal of Experimental Psychology* 32 (1): 3–25. doi: 10.1080/00335558008248231.
- Saxe, Rebecca, and Nancy Kanwisher. 2003. "People Thinking about Thinking People: The Role of the Temporo-Parietal Junction in 'Theory of Mind'." *NeuroImage* 19 (4): 1835–1842. doi: 10.1016/S1053-8119(03)00230-1.
- Saxe, Rebecca, and Lindsey J. Powell. 2006. "It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind." *Psychological Science* 17 (8): 692–699. doi: 10.1111/j.1467-9280.2006.01768.x.
- Scholz, Jonathan, Christina Triantafyllou, Susan Whitfield-Gabrieli, Emery N. Brown, and Rebecca Saxe. 2009. "Distinct Regions of Right Temporo-Parietal Junction Are Selective for Theory of Mind and Exogenous Attention." *PLoS ONE* 4 (3): e4869. doi: 10.1371/journal.pone.0004869.
- Somerville, Leah H., Rebecca M. Jones, Erika J. Ruberry, Jonathan P. Dyke, Gary Glover, and B. J. Casey. 2013. "The Medial Prefrontal Cortex and the Emergence of Self-Conscious Emotion in Adolescence." *Psychological Science* 24(8): 1554–1562. doi: 10.1177/0956797613475633.
- Van Horn, John Darrell, Andrei Irimia, Carinna M. Torgerson, Micah C. Chambers, Ron Kikinis, and Arthur W. Toga. 2012. "Mapping Connectivity Damage

in the Case of Phineas Gage.” *PLoS ONE* 7 (5): e37454. doi: 10.1371/journal.pone.0037454.
 Young, Liane, Joan Albert Camprodon, Marc Hauser, Alvaro Pascual-Leone, and Rebecca Saxe. 2010. “Disruption of

the Right Temporoparietal Junction with Transcranial Magnetic Stimulation Reduces the Role of Beliefs in Moral Judgments.” *PNAS* 107 (15): 6753–6758. doi: 10.1073/pnas.0914826107.

OTHER SOURCES USED

Dunne, Simon, and John P. O’Doherty. 2013. “Insights from the Application of Computational Neuroimaging to Social Neuroscience.” *Current Opinion in Neurobiology* 23 (3): 387–392. doi: 10.1016/j.conb.2013.02.007.
 Kobayashi, Shunsuke, and Wolfram Schultz. 2008. “Influence of Reward Delays on Responses of Dopamine Neurons.” *Journal of Neuroscience* 28 (31): 7837–7846. doi: 10.1523/JNEUROSCI.1600-08.2008.
 Passingham, Richard E., and Steven P. Wise. 2012. *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight*. Oxford: Oxford University Press.

Smith, Christopher T., and Charlotte A. Boettiger. 2012. “Age Modulates the Effect of COMT Genotype on Delay Discounting Behavior.” *Psychopharmacology* 222 (4): 609–617. doi: 10.1007/s00213-012-2653-9.
 Winstanley, Catharine A., Jeffrey W. Dalley, David E. H. Theobald, and Trevor W. Robbins. 2003. “Global 5-HT Depletion Attenuates the Ability of Amphetamine to Decrease Impulsive Choice on a Delay-Discounting Task in Rats.” *Psychopharmacology* 170 (3): 320–331. doi: 10.1007/s00213-003-1546-3.

FURTHER READING

Adolphs, Ralph. 2010. “Conceptual Challenges and Directions for Social Neuroscience.” *Neuron* 65 (6): 752–767. doi: 10.1016/j.neuron.2010.03.006.
One of the leading figures in social neuroscience considers the challenges confronting, and promises awaiting, this field.
 Amodio, David M. 2010. “Can Neuroscience Advance Social Psychological Theory? Social Neuroscience for the Behavioral Social Psychologist.” *Social Cognition* 28 (6): 695–716. doi: 10.1521/soco.2010.28.6.695.
A perspective on the value, or otherwise, of bringing neuroscience methods into the traditionally behavioral field of social psychology.
 Kennedy, Daniel P., and Ralph Adolphs. 2012. “The Social Brain in Psychiatric and Neurological Disorders.” *Trends in Cognitive Sciences* 16 (11): 559–572. doi: 10.1016/j.tics.2012.09.006.
Psychiatric and neurological disorders have informed the concept of the social brain, and the authors argue that the social brain, and its dysfunction and recovery, must be understood not in terms of specific structures, but rather in terms of their interaction in large-scale networks.
 Macmillan, Malcolm. 2000. *An Odd Kind of Fame: Stories of Phineas Gage*. Cambridge, MA: MIT Press.
Engrossing, thoughtful account of all things Gage, covering Gage’s personal history, cultural history, history of science, and implications of the case for cognitive neuroscience, including an emphasis on the localization–mass action debate that reigned over the decades surrounding the turn of the twentieth century.
 Seabrook, John. 2008. “Suffering Souls: The Search for the Roots of Psychopathy.” *New Yorker*, November 10. http://www.newyorker.com/reporting/2008/11/10/081110fa_fact_seabrook.
Interesting profile of cognitive neuroscientist and psychopathy researcher Kent Kiehl.
 Seeley, William W., Juan Zhou, and Eun-Joo Kim. 2012. “Frontotemporal Dementia: What Can the Behavioral Variant Teach Us about Human Brain Organization?” *Neuroscientist* 18 (4): 373–385. doi: 10.1177/1073858411410354.
This review provides entrée into a fascinating (ongoing) scientific detective story about the causes of this insidious disease, which prominently compromises social behavior, personal conduct, and decision making.

CHAPTER 17

EMOTION

KEY THEMES

- Neuropsychological and neuroimaging data suggest that the amygdala performs a kind of implicit, quick-and-dirty triage of visual stimuli, signaling, for example, whether or not faces look trustworthy, whether they belong to a certain race, or whether they are expressing an emotion. The evidence further suggests that the amygdala's "decision rules" are learned, and can therefore change with experience.
- Scientific interest in the amygdala was triggered by the drastic and bizarre changes in behavior associated with surgically produced Klüver–Bucy syndrome.
- Intensive study of Pavlovian fear conditioning has yielded detailed knowledge, at the circuit and molecular level, of the plastic change in the amygdala that underlies acquisition and expression of conditioned fear.
- The emotional component of declarative memories also seems to depend on the amygdala and, in particular, its influence on norepinephrine levels via its outputs to the locus coeruleus.
- From a neurobiological perspective, emotions may derive from the activity of systems related to survival, and perhaps to the underlying assessment that a stimulus or situation is safe to approach, or best to avoid.
- The extinction of a fear memory depends on new learning in the amygdala, but also on top-down input from ventromedial PFC (infralimbic (IL) cortex in the rat).
- Unlike conditioned memories, which are implicit and long-lasting, our current emotional state is amenable to real-time volitional control. The mechanisms by which we implement this control are the same as those used for the control of other domains of cognition.
- Many types of psychopathology (post-traumatic stress disorder, depression, anxiety disorder) are characterized by abnormal neural correlates of emotion regulation.

CONTENTS

KEY THEMES

EMOTION PROCESSING AND MENTAL HEALTH

TRUSTWORTHINESS REVISITED – AGAIN

A role for the amygdala in the processing of trustworthiness

Neuropsychological evidence

fMRI evidence

Implicit information processing by the amygdala

THE AMYGDALA

Klüver–Bucy syndrome

Pavlovian fear conditioning

Induction

Distinguishing avoidance learning from fear conditioning

Emotional content in declarative memories

Pharmacological interference with norepinephrine

Emotional declarative memory in patients with amygdala lesions

The amygdala's influence on other brain systems

What emotions does the amygdala process?

WHAT IS AN EMOTION?

THE CONTROL OF EMOTIONS

Extinction

Research in the rodent

The roles of midline frontal cortex

Fear conditioning and extinction in the human

“Real-time,” volitional control of emotions

The good and the bad of emotion regulation

THE AMYGDALA AS AFFECTIVE HUB

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

EMOTION PROCESSING AND MENTAL HEALTH

Depression and anxiety are not explicit emotional behaviors – they are, at their core, exaggerated implicit emotional reactions that have not been properly regulated. Identification of the genetic signatures that predict these exaggerated reactions, as well as the compromised prefrontal function that yields to these exaggerations, will clarify the predisposition of particular individuals to psychopathology. (Hariri and Whalen, 2011, p. 4)

This quote from affective neuroscientists Ahmad Hariri and Paul Whalen captures many aspects of the study of emotion that will be featured in this chapter: the neural circuitry underlying the processing of (“reactions to”) emotional stimuli; the control of this processing; the neural bases of the distinction between explicit and implicit processing of emotion; and the relevance of this work to mental health. The study of the neural bases of emotions – affective neuroscience – is a rich field worthy of its own textbook. Indeed, many such textbooks have been written. The goal of this chapter will be to convey some of the principles of affective neuroscience, and also to consider ways in which, as we have already seen with social context, emotional context can influence the cognitive processes to which this book is primarily devoted.

TRUSTWORTHINESS REVISITED – AGAIN

So it turns out that, almost as if conjured on command to meet the needs of this book’s author, *trustworthiness* has previously been studied in the context of affective neuroscience. And because the studies in question engage this social construct with which we’re already familiar, as well as a cognitive process – face perception – that we have also considered in depth, they offer a tailor-made opportunity for us to introduce a neural structure that will be front-and-center this chapter: the amygdala.

A role for the amygdala in the processing of trustworthiness

Neuropsychological evidence

A seminal study linking the amygdala to judgments of trustworthiness came from the team of Ralph Adolphs, Daniel Tranel, and Antonio Damasio, who reported on judgments of approachability and trustworthiness in three patients with bilateral amygdala lesions. The findings, illustrated in *Figure 17.1*, showed that the patients tended to judge all faces to be approachable and trustworthy, even those belonging to the 50 most negative faces, as rated by an independent group of individuals. Recall from *Chapter 10* that the amygdala is located in the medial temporal lobe, just rostral, medial, and superior to the rostral end of the hippocampus. Although it is in close anatomical proximity to ventral anterior IT cortex, it is not typically considered a part of the visual system (a point that is germane to the findings of this study, and many others to follow). Strikingly, however, these patients did not differ from controls when making the same judgments based on *verbal* descriptions of different individuals. Thus, the impairment seems tied to the visual modality. Based on this finding, Adolphs, Tranel, and Damasio (1998) proposed that “The amygdala appears necessary to trigger the retrieval of information on the basis of prior social experience or innate bias in regard to certain classes of faces” (p. 472). By this account, the social information itself is stored elsewhere, and accessible via other channels, such as spoken language.

fMRI evidence

Several fMRI studies prompted by the Adolphs, Tranel, and Damasio (1998) finding have shown that the magnitude of amygdala responses to faces corresponds to judgments of the trustworthiness of each face. Remarkably, one such study has shown that the correlation is stronger between the amygdala of an individual and the group-average ratings of trustworthiness made by 129 unscanned subjects than it is for that individual’s own ratings of trustworthiness (Engell, Haxby, and Todorov, 2007; *Figure 17.2*)! I know; it also took me a while to wrap my brain around this finding. How could a person’s brain show a response that corresponds more closely to a

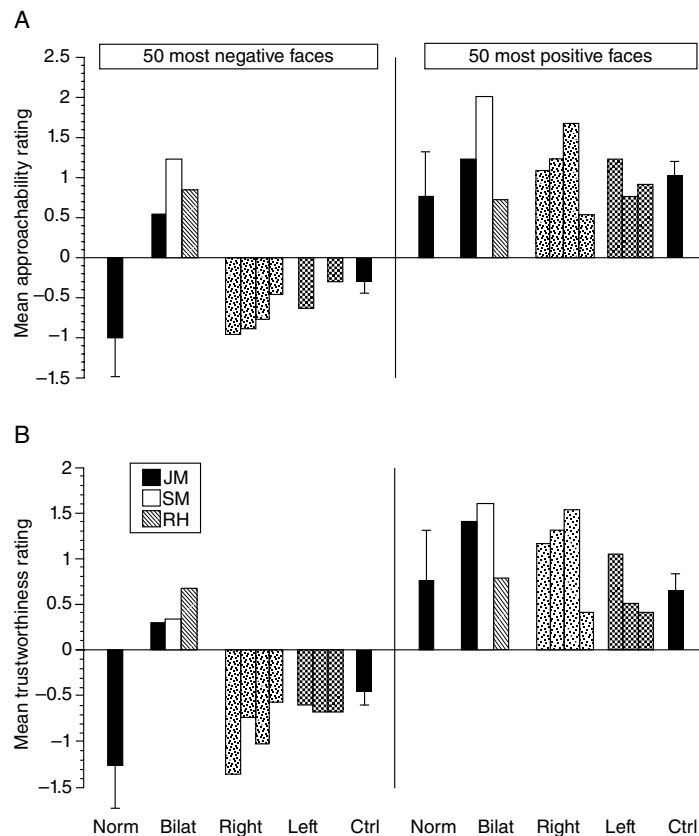


FIGURE 17.1 Performance of three patients with bilateral amygdala lesions (Bilat) – relative to neurologically healthy control subjects (Norm), unilateral right or left amygdala lesion patients, and non-amygdala lesion control subjects (Ctrl) – on rating the approachability and trustworthiness of 50 visually presented faces rated “most negative,” and 50 “most positive,” by an independent group. Source: Adolphs, Tranel, and Damasio, 1998. Reproduced with permission of Nature Publishing Group.

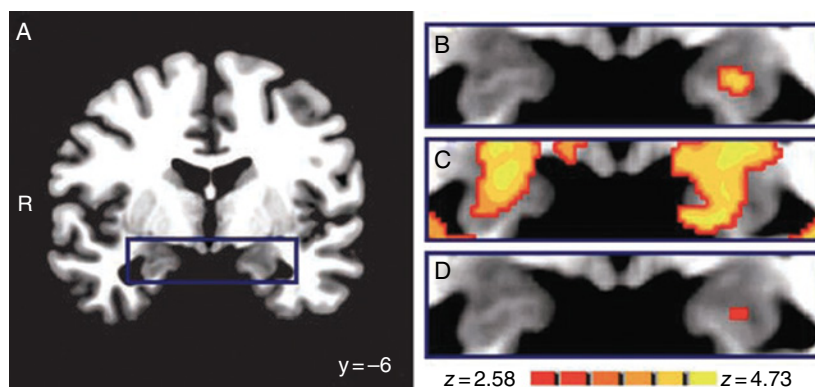


FIGURE 17.2 Amygdala responses to visually presented faces correlate with “consensus” judgments of trustworthiness. **A.** Coronal section with a rectangle indicating the area that is magnified in B–D. The black regions immediately lateral to each amygdala are the temporal horns of the lateral ventricles. **B.** The correlation of individual judgments of trustworthiness (acquired in the post-scanning test) with amygdala activity. **C.** Correlation of “consensus” judgments of trustworthiness, the average across 129 subjects who were not scanned, with amygdala activity. **D.** Correlation with individual judgments after partialling out variance shared between individual and consensus judgments. Source: Engell, Haxby, and Todorov, 2007. Reproduced with permission of The MIT Press.

group-average judgment than to that own person's judgment? The answer to this question requires us to keep in mind that this finding only pertained to one region within this person's brain. As Engell, Haxby, and Todorov (2007) note, their finding is "consistent with the notion that the amygdala is involved in the initial automatic assessments of trustworthiness based on facial features, but that other neural systems, which are more closely linked to idiosyncratic perception, modulate the amygdala's influence on behavior" (p. 1516).

A final note about this study is that Engell, Haxby, and Todorov (2007) told their subjects that they were participating in a face memory experiment, and made no mention of trustworthiness prior to the scanning session. Then, with the same logic as a subsequent-memory study (*Chapter 11*), the researchers measured behavioral correlates of the actual mental construct of interest, in this case trustworthiness judgments, after the scanning session. This means that the activity measured in their scans corresponded to what could be characterized as "implicit trustworthiness judgments." That is, the subjects presumably weren't thinking to themselves *Is this person trustworthy?* as they viewed each face; they were just trying to memorize them. Nonetheless, a part of their brain was responding to each face in a way that very strongly mirrored how they would later judge trustworthiness when asked to do so explicitly.

If we relate the findings and interpretation of the Engell, Haxby, and Todorov (2007) study to those of Adolphs, Tranel, and Damasio (1998), a decidedly visual role for the amygdala seems to emerge. Recall that the patients with bilateral amygdala lesions were not impaired at making approachability and trustworthiness judgments when they listened to brief spoken biographies of hypothetical individuals. Thus, despite what we earlier said about the amygdala not being considered a part of the visual system, the operation performed here seems to be a fundamentally visuo-perceptual one: it may be that the amygdala detects configurations of physical properties that can be used to classify other people according to any number of categories, including their moral character, and the emotion that they are currently experiencing. By this view, the amygdala may carry out fast, unconscious, "triage" operations on incoming sensory information, upon which this initial classification, this "early-warning

signal," is fed to relevant cortical areas that can then engage in the more deliberative processing that we detailed in *Chapter 15*.

Implicit information processing by the amygdala

One question for us to keep in mind as we proceed through this chapter is whether this function of the amygdala is limited to vision. For example, what if one presented to subjects the voices of different individuals, all speaking the same sentence, such that physical properties like timbre and emotional inflection, but not the semantic content of the speech, would vary as a function of moral character? Would the amygdala also implicitly classify this auditory information as a function of trustworthiness? We'll get a sense of the modality specificity of the amygdala in the next section, when we consider its functional anatomy in detail. But before leaving our friend *trustworthiness*, another important question to ask is whether this phenomenon of implicit social "judgment," of unconscious classification of people that we encounter, generalizes to other dimensions along which one can evaluate other people. The answer, as explored in *Real-World Applications 17.1*, seems to be "yes."

We will return to the study of amygdala processing of implicit bias later in this chapter, when we consider the volitional control of emotion. At present, however, we will pursue a different theme highlighted in *Real-World Applications 17.1*, which is that the principles of plasticity and learning in the amygdala seem to play a prominent role in this region's functions. Indeed, we've reached a point where, having linked a handful of intriguing findings to the activity of this neural structure, we really need to learn more about the structure itself – its afferents, its internal anatomy and physiology, and its efferents.

THE AMYGDALA

Klüver–Bucy syndrome

The idea that the amygdala plays a central role in emotion processing has its roots in the work of German-born, US-based experimental psychologist

REAL-WORLD APPLICATIONS

17.1 Neural bases of racial stereotyping

The implicit association test (IAT) is widely used by social psychologists to measure implicit social judgments, such as the judgments of trustworthiness illustrated in *Figure 17.1* and *Figure 17.2*. A study of racial bias by affective neuroscientist Elizabeth Phelps and colleagues (2000) used a design similar to that of Engell, Haxby, and Todorov (2007) – first scan, then obtain behavioral measures of the subjects' evaluation of the stimuli and see if the latter relate to the former – with the important difference that the post-scan behavioral measures included a test of implicit cognition, in addition to a test of explicit cognition.

Research subjects in experiment 1 of Phelps et al. (2000) were White Americans who were told they were participating in a memory study. During scanning, they viewed photographs of nine White male faces and nine Black male faces. After the scan, subjects were first administered an IAT during which they made interleaved speeded judgments of whether a face was Black or White (race classification) and whether a word was good or bad (word classification). The logic of the IAT is to vary the classification-to-response-button mapping, such that on some blocks the race classification maps Black to left and White to right, and on others the mapping is reversed. Independently, the word-classification mapping to the same two buttons is also varied, such that one can then compare the RT for endorsing a good word as “good” when the same button was required for classifying a face as Black, vs. the RT for endorsing a good word as “good” when that button was required for classifying a face as White. The oft-replicated finding is that White Americans show an implicit bias, in that they are slower to endorse good words as “good” when the same button is associated with Black faces. Notably, this effect is statistically independent of one's professed, explicit stance with regard to racial equality.

Experiment 1 of Phelps et al. (2000) did, indeed, replicate the standard finding of an implicit negative association between Black faces and positive words. This was despite the fact that the subjects also were rated with a mean of 1.89 on the Modern Racism Scale, a test of explicit racial attitudes that was administered after the IAT. The scale ranges from 1 (strongly pro-Black) to 6 (strongly anti-Black). Comparison of the behavioral vs. fMRI data showed that the magnitude of implicit anti-Black bias in an individual's IAT score was positively correlated with the extent to which the amygdala response in that subject was greater for Black than White faces. There was no evidence for a similar relationship between performance on the Modern Racism Scale and amygdala responsivity.

In experiment 2, Phelps et al. (2000) repeated their experimental procedure, but with the faces of famous and positively regarded Black and White men. For these stimuli, the IAT revealed diminished evidence for differential association of Black vs. White faces with positive words, and these scores no longer correlated with amygdala responses. Because experiment 1 used unfamiliar faces and experiment 2 famous faces, these findings were interpreted as evidence that the fast, implicit social categorization performed by the amygdala is learned, rather than hardwired. (Find out more about the IAT, and test yourself if you like, at [Web Link 17.1.](#))

Heinrich Klüver (1897–1979) and American neurosurgeon Paul Bucy (1904–1992). Klüver had pursued the first three years of his doctoral research at the University of Hamburg under the mentorship of Max Wertheimer (1880–1943), a founder of the tradition of **gestalt psychology**. After immigrating to the United States in 1923, he received his PhD from Stanford University in 1924 for research on visual imagery (“eidetic vision”) in children (presumably getting credit for work done in Hamburg). Soon thereafter he began to compare his findings in children with subjective experiences produced by consuming the dried tops of the cactus *Lophophorus Williamsii* (a.k.a., peyote), which contains the psychoactive compound mescaline (*Tip 17.1*). These early studies of self-experimentation led to subsequent studies of the effects of mescaline on the behavior of monkeys, and this led Klüver to Paul Bucy.

Of particular interest to Klüver was the observation that mescaline produced chewing and licking behaviors, as well seizures, in monkeys, and these suggested a type of seizure that the neurologist John Hughlings Jackson (*Chapter 1*) had named “uncinate fits.” These seizures, implicating the temporal-lobe terminations of the uncinate fasciculus (*Figure 16.3*), are characterized by hallucinations of taste and odor, no doubt a consequence of this region’s involvement in olfactory processing. Klüver’s goal, therefore, was to

determine whether lesions of the medial temporal lobes would produce comparable effects, thereby localizing the neural locus of mescaline’s psychoactive effects. He enlisted Bucy, a neurosurgeon, to perform the surgical lesions. Although Klüver wanted Bucy to restrict the region to the uncus (adjacent to the piriform cortex, see *Chapter 16*), Bucy “felt he couldn’t perform such an operation and decided to take out the whole temporal lobe” (Nahm and Pribram, 1998, p. 9). The syndrome that resulted, which was exacerbated by an equivalent surgical removal in the other hemisphere, was dramatic, and characterized by five cardinal symptoms: (1) “psychic blindness” (later to be known as associative agnosia (*Chapter 9*)); (2) “tameness” (i.e., docility); (3) hyperorality, characterized by putting nearly everything in the mouth to “explore” it (this has been reported to include such obviously hazardous items as lighted cigarettes); (4) dietary changes, including hyperphagia (an extreme drive to consume food); and (5) hypersexuality, including the mounting of animals from different species. Anecdotal reports also indicated that the monkeys lost their fear of snakes and of humans, the former of which is believed to be innate.

In the 1950s, research by Lawrence Weiskrantz (later to mentor Charles Gross (*Chapter 1* and *Chapter 5*), and of “Warrington and Weiskrantz” renown (*Thought Box 11.1*)) localized most of the symptoms of the Klüver–Bucy syndrome to the amygdala, and established the idea that this structure was important for processing the rewarding or punishing value of stimuli. This set the stage for enormously influential breakthroughs from studies of the role of the amygdala in Pavlovian fear conditioning.

Tip 17.1

Klüver’s 1926 publication “Mescal Visions and Eidetic Vision,” and his 1928 monograph *Mescal: The “Divine” Plant and Its Psychological Effects*, not only predated, but, by emphasizing alterations of the perceptual experience, also foreshadowed the influence that mescaline would have on literary culture of the 1950s and 1960s. Those who consumed and wrote about, and/or under the influence of, mescaline, include Aldous Huxley (*The Doors of Perception*, 1954), Ken Kesey, Carlos Castaneda, Hunter S. Thompson, and Jean-Paul Sartre. Mescaline is a 5-HT_{2A} agonist, the presumed basis for its psychoactive effects.

Pavlovian fear conditioning

Induction

Recall from *Chapter 10* that Pavlovian conditioning entails pairing an unconditioned stimulus (US) with a conditioned stimulus (CS), in temporal contiguity, until an association between the two is learned, at which time the CS, alone, will produce the conditioned response. In the classic fear conditioning protocol the CS is an auditory tone and the US is a footshock (see *Real-World Applications 10.1*; *Tip 17.2*). Decades of intensive research, beginning in the 1980s,

have established in exquisite detail the mechanisms underlying fear conditioning in the mammal. For the following summary to make sense, you should first familiarize yourself with the circuitry of the amygdala, and of its afferents and efferents, by studying *Figure 17.3*.

Tip 17.2

This is such a commonly used protocol that “footshock” has become an accepted word in this literature.

Okay, now that you’ve familiarized yourself with LeDoux’s diagram, here’s his summary of fear conditioning. Fasten your seatbelt:

CS and US convergence occurs in the lateral nucleus of the amygdala (LA), and specifically in the dorsal subregion of the LA. This convergence leads to synaptic plasticity and the formation of a CS–US association. Damage to LA, inactivation of LA, or manipulation of a variety of molecular pathways in LA prevents fear conditioning. A second important region is the central nucleus of the amygdala (CE). Manipulations of this region also disrupt conditioning. LA and CE are connected directly and by way of various intra-amygdala pathways. Once the CS–US

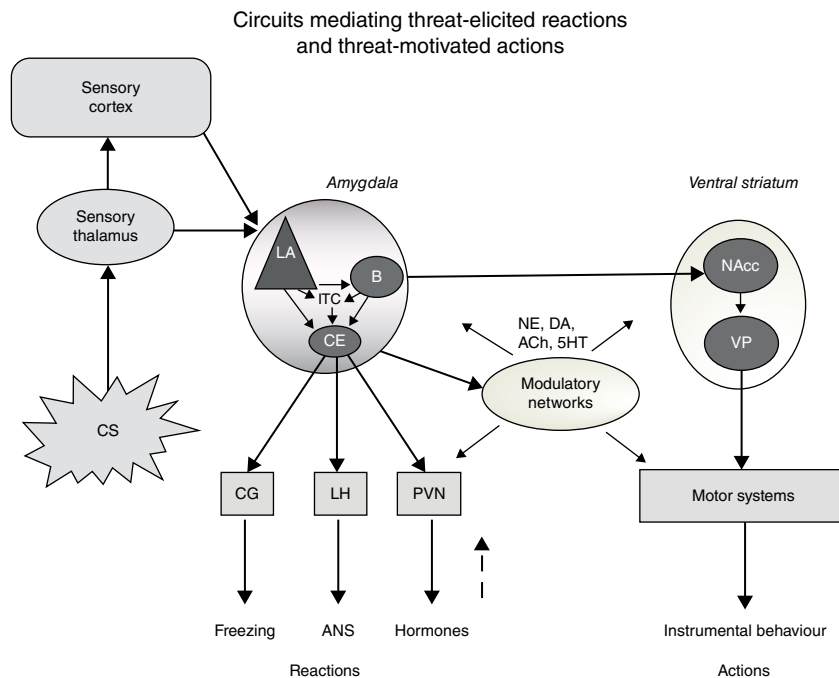


FIGURE 17.3 A schematic diagram of what is sometimes called the “fear circuit,” which LeDoux (2012a) frames here as the threat-processing circuit. The CS, which is most often a tone or a light, is first registered in the sensory thalamus (e.g., the MGN or LGN) then conveyed to the lateral nucleus of the amygdala (LA) via two routes, the putative direct “low road” and the cortical “high road.” (The former presumably carries only low-resolution, magnocellular-dominated signals, but is central to many models of amygdala function; some, however, dispute its existence.) Within the amygdala complex are the lateral (LA), basal (B), and central (CE) nuclei, as well as the GABAergic intercalated cells (ITC). The CE sends outputs to regions that produce the body’s internal reactions to fear: the central gray (CG); the lateral hypothalamus (LH; which in turn activates the autonomic nervous systems (ANS)); and the paraventricular nucleus (PVN) of the hypothalamus. The basal nucleus sends outputs to the nucleus accumbens (NAcc) and ventral pallidum (VP) of the basal ganglia, thereby influencing motor actions. “Modulatory networks” refers to subcortical nuclei that innervate the brain to deliver the modulatory neurotransmitters listed. Source: LeDoux, 2012a. Reproduced with permission of Elsevier.

association is formed, later exposure to the CS results in the retrieval of the learned association formed by CS–US convergence during conditioning. Information then flows from LA to CE, which then connects to hypothalamic and brainstem areas that control behavioral, autonomic, and hormonal responses that help the organism cope with the threat. Plasticity also occurs in CE, and in CS processing regions and motor control regions. (LeDoux, 2012a, p. 435)

As noted by LeDoux, this is a simplified description that omits many details. Should one want more detail, two places to begin are LeDoux (2000) and Davis and Whalen (2001), both listed in the *Further Reading* section. Here, we will limit ourselves to just some key concepts.

First, the necessary locus of LTP, without which fear conditioning does not occur, is the LA nucleus. Among the manipulations of LA that block fear conditioning is the administration of protein synthesis inhibitors, which block the formation of new synapses within the LA and between the LA and other amygdala nuclei. This is an evolutionarily ancient form of learning, observable in many invertebrate organisms with much simpler nervous systems. (These animals have neither amygdalae, nor LA nuclei. However, synaptic plasticity underlying Pavlovian conditioning in these animals results from very similar, if not identical, intracellular molecular cascades that are triggered by the network-level convergence of signals relating to the CS and the US.)

Second, downstream from synaptic changes in LA, (at least) two different kinds of learning are supported by (at least) two pathways in the amygdala. Most of our attention in this chapter will focus on Pavlovian fear conditioning, which is mediated by circuits connecting LA to CE, and expressed via CE's outputs to systems that implement the innate "Reactions" (bottom-left of *Figure 17.3*). A second pathway, from LA to the basal nucleus, implements a different kind of learning, **avoidance learning**.

Distinguishing avoidance learning from fear conditioning

Avoidance learning refers to context-dependent skeletomotor behaviors that, when engaged, enable the organism to avoid the adverse outcome predicted by

the CS. One laboratory example of an avoidance behavior would be if the floor of the testing chamber has two different textures, and the rat learns to avoid a shock by, each time a tone is played, moving to the "safe zone" in the chamber. An avoidance behavior from everyday life is illustrated during my daily walk home from work. Having been unpleasantly startled more than once by the aggressive dog who is sometimes outdoors in front of a particular house, I have learned to turn left at the intersection preceding this house so as to avoid the unpleasant encounter.

There are three points to emphasize about the amygdala-mediated "instrumental behavior" (*Figure 17.3*, lower right) of which avoidance learning is a subtype. First is that it entails engaging the motor system via the basal nucleus-to-ventral striatum efferent pathway of the amygdala (i.e., a different pathway from LA to CE). Second is its context-dependence. If I'm on a sailboat on Lake Mendota and suddenly I see a large motor boat bearing down on me, the notion of *turning left at the intersection of Kendall St. and Prospect Ave. to avoid this threat* doesn't even make sense. That learned behavior is only relevant in the context of avoiding the dog on the walk home from work. Third is that the behaviors acquired via avoidance learning aren't emotional per se. Returning again to my pedestrian commute, if, on my way home from work, I need to pick up my son at his friend Alex's house I would also turn left at that same intersection, because that's the way to get to Alex's house. This is in contradistinction to fear conditioning, in which the CS triggers the same innate, hardwired defense mechanisms as does a US. Thus, for example, when a dog suddenly barks and lunges (a US), one's first responses are to freeze and to activate the autonomic nervous system (i.e., the "fight or flight" reflex). Similarly, when the sailor suddenly sees the large motor boat bearing down on her (also a US), her first response is also to freeze and to activate the autonomic nervous system. (It is only "a split second" after this initial stereotyped response that she engages the contextually appropriate response of grabbing the tiller to steer the sailboat out of the path of the oncoming motor boat.)

Emotional content in declarative memories

Pharmacological interference with norepinephrine

The two types of memory summarized in the preceding section, fear conditioning and avoidance learning, are nondeclarative – they do not depend on the integrity of the hippocampus or rhinal cortices. However, declarative memories can also have an emotional component. Is the amygdala involved in these? Before answering this question, we'll first revisit the study of pharmacological interventions with emotional declarative memory. Recall from *Real-World Applications 10.1* that administration of the β -adrenergic antagonist propranolol at the time of retrieval of an emotional memory “knocks down” the emotional content of the memory on subsequent retrievals. This finding provided evidence that reconsolidation under normal circumstances entails the activation of the subcortical “modulatory networks” illustrated in *Figure 17.3*. Those reconsolidation studies considered in *Chapter 10* built on the earlier work performed by neurobiologists Larry Cahill, James McGaugh, and colleagues, as summarized here. In one celebrated study, Cahill, Prins, Weber, and McGaugh (1994) used 12 images to construct a story with two variants, one emotionally arousing and one emotionally neutral. Both involved a mother and son leaving home, a hospital emergency room, and hailing a taxi. In the neutral story, the mother and son walk to the hospital where the father works, passing automobile wreckage in a junkyard along the way. They observe an emergency preparedness drill (which includes a brain scan and makeup artists creating simulated injuries), and the mother phones the day care facility where her younger child is enrolled, then hails a taxi to go pick up this child. The arousing story begins the same way, but then the boy is critically injured by a collision between cars while he and the mother are crossing the street, the brain scan and gore at the hospital are described as medical procedures undertaken to save the boy's life, and the phone call and taxi are the same.

For their experiment, Cahill et al. (1994) divided their subjects into four groups. An hour prior to hearing the story, subjects in two of the groups were

administered propranolol, and those in the other two were administered a placebo. Subsequently, subjects in one of the two propranolol groups and in one of the two placebo groups heard the arousing version of the story, and those in the remaining two groups heard the neutral version. On the surprise memory test administered one week later, the *arousing-story/placebo* group (A/P in *Figure 17.4*) showed the expected benefit that emotional content gives to memory performance. Such emotion-related enhancement was not present, however, for the *arousing-story/propranolol* (A/BB) group, whose memory performance for the middle part of the story didn't differ from that of the two neutral-story groups (*Figure 17.4.A*). Notably, however, the A/BB did not differ from the A/P group with regard to how emotional they found the story to be, when asked to rate it immediately after having heard it. This latter finding suggests that the effect of propranolol was specific to emotional memory encoding and had minimal, if any, effect on the experience of emotion at the time that the story was presented.

Emotional declarative memory in patients with amygdala lesions

Now, you may be wondering why the description of the Cahill et al. (1994) experiment is presented in this section of the chapter that is focused on the amygdala. After all, propranolol administration in this study was systemic, and so antagonism of β -adrenergic receptors was likely to have occurred throughout the brain, as well as in the periphery. The reason for including it in this section is because a subsequent study used the same experimental materials and procedure to test the emotional memory of patients with bilateral amygdala lesions. In it, a team led by Ralph Adolphs, and including Larry Cahill, found that the performance of amygdala-lesioned patients was very similar to that of healthy individuals who had been administered propranolol: first, just like the A/BB subjects in *Figure 17.4.A*, their declarative memory was selectively impaired for emotional portions of the story; second, like the A/BB subjects in *Figure 17.4.B*, their immediate rating of the emotionality of the story was comparable to that of control subjects(!). This certainly came as a surprise to me. Does this mean that amygdala isn't

FIGURE 17.4 Effect of propranolol on emotional memory encoding. Source: Cahill, Prins, Weber, and McGaugh, 1994. Reproduced with permission of Nature Publishing Group.

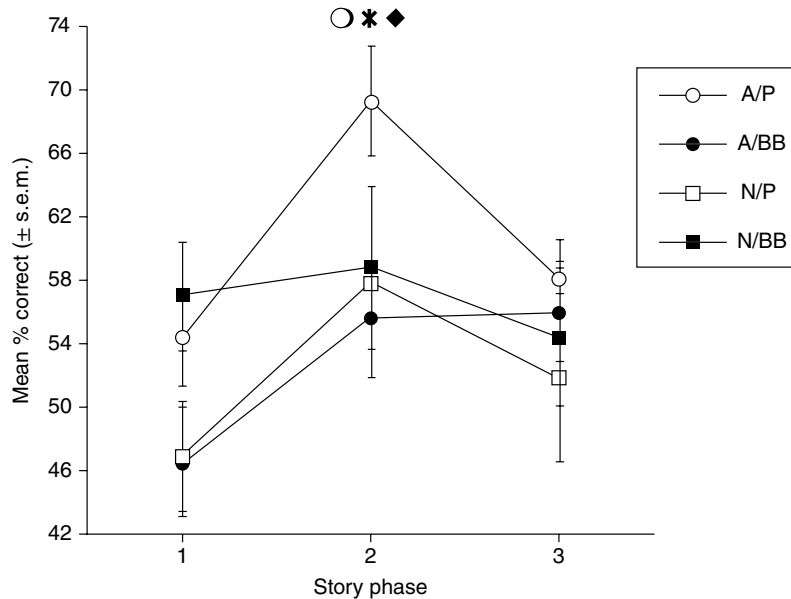


FIGURE 17.4.A Performance of four groups of subjects on a surprise memory test administered one week after hearing either the emotionally arousing (A) or neutral (N) version of a story, prior to which they had been administered either a placebo (P) or the drug propranolol (a “beta blocker” (BB)). These factors were fully crossed in a 2 x 2 between-subjects design. Story phases 1 and 3 were identical. The A/P group, but not the A/BB group, demonstrated superior memory performance for story phase 2, which contained the emotionally arousing content. The A/BB group’s performance did not differ from that of the groups that heard the emotionally neutral version of the story.

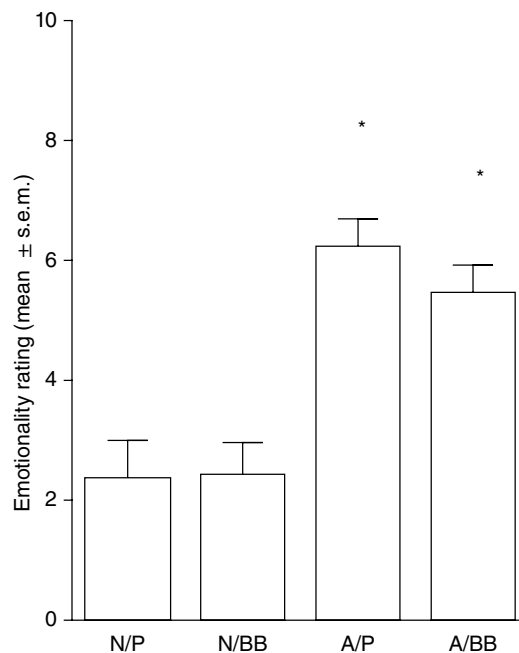


FIGURE 17.4.B Immediately after hearing the story, subjects were asked to rate on a 1–10 scale how emotional they found the story. Both groups that had heard the emotionally arousing version of the story rated it to be more emotional than did the groups that had heard the emotionally neutral version.

necessary for the experience of emotion?!? Anecdotaly, Adolphs, Cahill, Schul, and Babinsky (1997) note that one of the two patients

explained that she had found the slide with the surgically repaired legs quite emotional and unpleasant, a comment she also made spontaneously when she first saw the slides. She has made similarly appropriate comments in regard to her experience of a large number of emotionally charged stimuli we have shown her, as well as in regard to emotional events from her own life.

Then, they go on to conclude that

It may thus be that the experience of emotional events does not depend essentially on the amygdala. The amygdala's function in humans appears to be rather more subtle, serving to mediate between the evaluation and experience of emotional events, on the one hand, and their acquisition and consolidation into long-term memory, on the other. (p. 298)

Other findings, from other groups, are consistent with this idea of a distinction between the experience of emotion and the incorporation of emotional content into memory. For example, a PET study by Cahill and colleagues (1996) suggested that activity in the amygdala correlated more strongly with memory for emotionally arousing stimuli than with subjects' subjective reports of their state of emotional arousal while in the scanner. Thus, we may need to look beyond the amygdala to get a fuller picture of how emotion is represented while we experience it "in real-time," and how it influences ongoing cognition and behavior. In so doing, we need keep in mind, however, the dictum of which we've periodically been reminded, which is that one's introspection is a notoriously poor indicator of what processes are actually unfolding between the ears. To cite just one example, recall from *Chapter 15* that our brain can decide when we will make a "volitional" response, and what that response will be, several seconds before we become aware of that decision. And just earlier in this chapter we've reviewed evidence of the unconscious influence of the amygdala on decisions related to trustworthiness and to racial bias. Thus, one caveat to keep in mind is whether, despite their reports on a single subjective measure of emotionality, patients with bilateral amygdala damage are truly experiencing the full richness of

the emotional response to an evocative story that control subjects are experiencing. The same holds for subjects with propranolol in their systems.

And so, to recap, our understanding at this point in the chapter is that the amygdala can influence our decisions, and possibly our emotions, implicitly. Additionally, however, there's reason to think that our explicit, subjective experience of emotions "in the moment" is supported, at least in part, by systems outside of the amygdala proper.

The amygdala's influence on other brain systems

As a transition toward considering the processing of emotion outside of the amygdala, we can start by considering how the amygdala influences other brain systems. For example, one possible explanation of the findings from the emotional memory studies that we have just reviewed (Cahill et al., 1994, and Adolphs et al., 1997) is that the amygdala's activation of the locus coeruleus (one of the "Modulatory networks" in *Figure 17.3*), and the consequent increase of norepinephrine levels in many parts of the brain, results in enhanced memory consolidation. If β -adrenergic receptors are blocked by a pharmacological agent, then the emotional memory effect is blocked. Similarly, if there is no amygdala to trigger the increased release of norepinephrine, then enhanced consolidation will not take place. Note that, unlike the midbrain dopaminergic nuclei A8–A10, whose output is focused on the basal ganglia and PFC, the locus coeruleus sends projections diffusely throughout the cortical mantle and to several subcortical targets. Further, β -adrenergic receptors are found throughout the brain, most densely in hippocampus, but also in cerebellum, basal ganglia, thalamus, midbrain, and cerebral cortex. Thus, it is difficult to know how anatomically localized or diffuse are the effects of propranolol that blocks the consolidation of memory for the emotional components of a story.

Consistent with the caveat about the interpretation of self-reports of the subjective experience of emotion, there is fMRI evidence that damage to the amygdala does, indeed, influence the processing of emotional stimuli. As illustrated in *Figure 17.5*, the fMRI response of the midfusiform gyrus to the visual

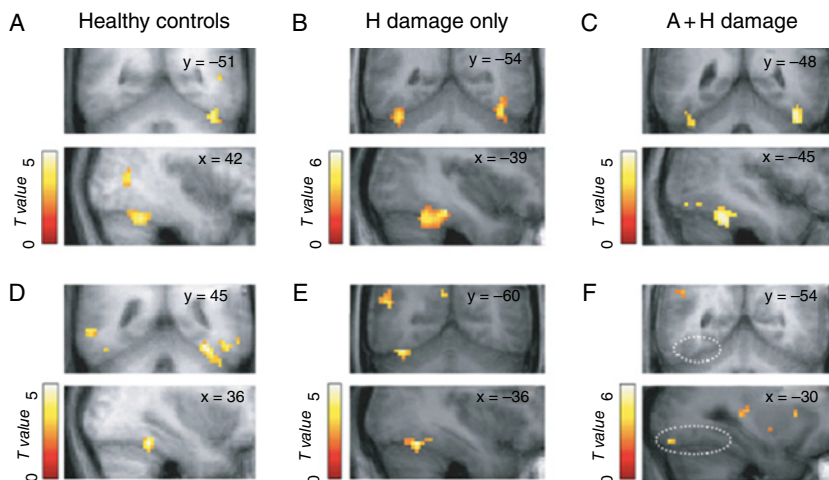


FIGURE 17.5 Activity evoked by attention to faces vs. houses, in control subjects, patients with medial temporal lobe damage encompassing the hippocampus, but not the amygdala (“H damage only”), and in patients with damage encompassing the hippocampus and amygdala (“A+H damage”). On each trial, two faces and two houses were presented, and, depending on the attentional cue, subjects had to decide whether either the two faces or the two houses were the same or different. Orthogonal to the factor of attention was the fact that half of the trials presented neutral faces and half presented fearful faces. **A–C** show that the response in fusiform gyrus for attention to faces was comparable across the three groups. **D–F**, in contrast, show a selective reduction of activity to fearful vs. neutral faces, regardless of whether subjects were attending to faces or houses, for the A+H group. Source: Vuilleumier, Richardson, Armony, Driver, and Dolan, 2004. Reproduced with permission of Nature Publishing Group

presentation of faces, relative to houses, is no different in patients with amygdala damage relative to control subjects. Where these patients differ, however, is in the comparison of the response to fearful vs. neutral faces. In the inferior temporal and occipital regions shown in *Figure 17.5*, as well as in primary visual cortex and superior temporal sulcus (not shown), patients with amygdala lesions do not show the robust increase in activity for fearful vs. neutral faces that is seen in healthy control subjects, as well as in patients whose lesions compromise the hippocampus but spare the amygdala. And so our understanding of the functions of the amygdala does, indeed, need to be nuanced.

What emotions does the amygdala process?

Before leaving the amygdala, let’s consider this one additional question. Up to this point, we’ve considered trustworthiness, we’ve considered aversive electric shock,

we’ve considered a story about a gruesome automobile vs. pedestrian accident, and we’ve considered faces expressing fear. One could make an argument that each of these can be reduced to the elemental emotion of fear. Another perspective that can account for these findings, however, is that the principal function of the amygdala is vigilance – serving as an early warning signal for any type of stimulus in the environment that might be predictive of a threat, whether the stimulus in question is inherently threatening or not. This is explored in *Thought Box 17.1*. (Additionally, *Web Clip 17.1* illustrates that the amygdala isn’t restricted to processing “negative” behaviors and emotions.)

WHAT IS AN EMOTION?

In the first part of this chapter, we have given considerable attention to how sensory information, such as a visually perceived face, gets classified, automatically

THOUGHT BOX

17.1 The whites of their eyes

A face expressing the emotion of fear isn't inherently frightening – someone who is really scared is unlikely to attack or otherwise threaten you directly. So why does the amygdala respond so potently to fearful faces? One idea is that it does so because its function is to detect potential sources of threat in the environment. The reasoning is that whatever it is that is scaring the person who you are looking at may represent a threat to you as well, and the sooner you determine what that threat is, the more likely it is that you can take the appropriate action to protect yourself. To be concrete, let's engage in mental time travel and visit our early-hominid ancestors at the proverbial watering hole. My ancestral grandmother is bending down to get a drink and, unbeknownst to her, a lion is sneaking up behind her and about to pounce. Although my ancestral grandmother doesn't see the lion, she does see the expression of fear on the face of *your* ancestral grandmother, who is facing the other direction and looking directly at the lion. That expression of fear is, in effect, a CS that triggers a conditioned fear response in my ancestral grandmother, and allows her to take evasive action that saves her life. (By the way: Thanks, I owe your ancestral grandmother one.)

To test this vigilance idea, affective neuroscientist Paul Whalen and colleagues (2004) devised an extreme test of amygdala responsivity. Whalen and others had already shown in previous studies that the amygdala responds more strongly to fearful than to happy faces, even when these are presented only very briefly (say, for 33 ms) and followed immediately by a neutral face that masks the emotional one, such that subjects have no subjective awareness of having seen a facial expression. Here we'll note that, although the vigilance hypothesis doesn't depend on the existence of a magnocellular "low road" projecting directly from thalamus to the amygdala (*Figure 17.3*), most assume it. And because faces are complex stimuli, it's unlikely that such a pathway would carry the kind of detailed, high-spatial-resolution information that would be needed, for example, to recognize a face. Thus, what if, Whalen asked, one were to take away all information related to face identity, and only give the amygdala the effectively binary information of whether or not the subject is in a "wide-eyed" state? The emotional states anger, fear, and surprise are associated with wide eyes, whereas happy, sad, and neutral emotional states are not. Would only this highly impoverished stimulus, varying only the amount of the sclera (i.e., the "whites" of the eyes) that is visible, be sufficient to differentially activate the amygdala? *Figure 17.6* illustrates that, remarkably, it is.

Subsequent work from Whalen's group has shown that the same can be shown for another low-level cue – pupil size. Indeed, not only is pupil dilation a visible signal of a heightened state of vigilance, research in humans and nonhuman animals has shown that direct electrical stimulation of the amygdala produces pupil dilation. Thus, when we meet on the street, your amygdala might be reading the state of my amygdala by gauging the diameter of my pupils!

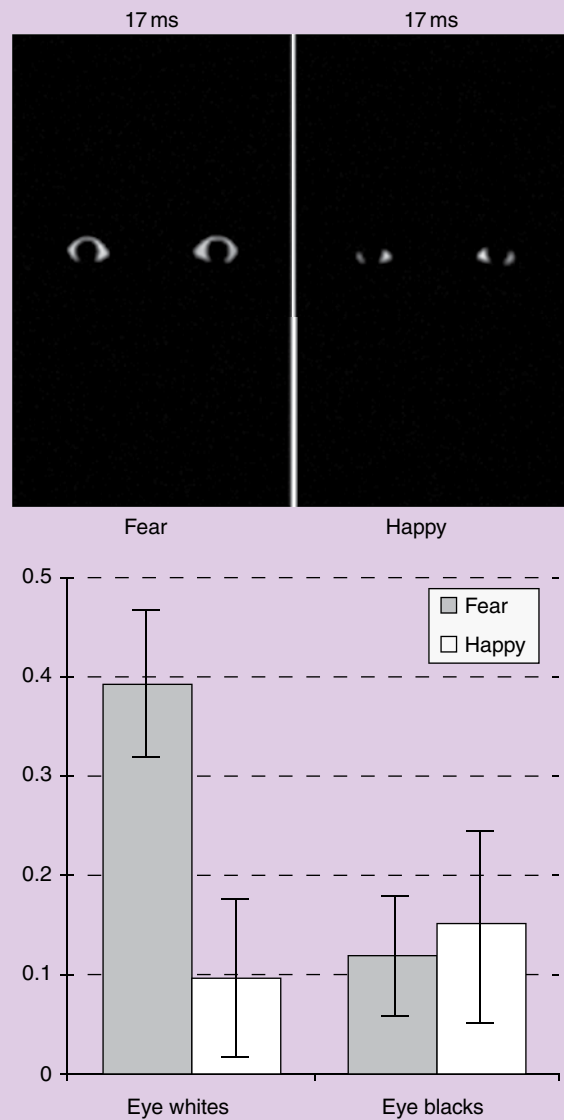
Thought Box 17.1 *Continued*

FIGURE 17.6 Differential amygdala response to fearful eyes vs. happy eyes. On the left are example stimuli from the two categories; on the right, the mean response in the ventral amygdala in the left hemisphere. “Eye blacks” refers to the control condition in which figure and ground luminances were reversed, such that the shapes of the sclera appeared as small black shapes on a white background. Although identical in shape, eye blacks never occur in nature, and so would not be expected to be detected as a threat-related CS. Source: Whalen, Kagan, Cook, Davis, Kim, Polis, McLaren, Somerville, McLean, Maxwell, and Johnstone, 2004. Reproduced with permission of AAAS.

and implicitly, along many dimensions. The outcome of these social classifications, depending on the context, can influence our behavior, as well as our subjective mood. Telltale signs of fear or vigilance on the part of others may activate our own autonomic nervous system, and perhaps heighten the level of circulating stress hormones, thereby making us feel nervous or uneasy. Unconscious biases relating to a person from a certain racial or ethnic group may influence how we feel about that person and how we interact with him or her. Thus, our interpretation of implicit social cues, and implicit social judgments that we make, can influence our emotions.

Another perspective from which we can understand the neural bases of emotion comes from psychology, and that's the distinction between approach vs. withdrawal behaviors. As framed by Richard Davidson and then-graduate student Wil Irwin (1999),

The approach system facilitates appetitive behavior and generates certain types of positive affect that are approach-related, for example, enthusiasm, pride, etc. ... This form of positive affect is usually generated in the context of moving toward a desired goal ... a second neural system [is] concerned with the implementation of withdrawal. This system facilitates the withdrawal of an individual from sources of aversive stimulation and generates certain forms of negative affect that are withdrawal-related. For example, both fear and disgust are associated with increasing the distance between the organism and a source of aversive stimulation. (pp. 12–13)

Davidson and colleagues have associated positive, approach-related emotions with elevated activity in the left PFC and negative, withdrawal-related emotions to elevated activity in the right PFC. This is seen in mood-induction studies, in which subjects who are made to feel happy or sad reliably produce greater activity in the left or right PFC, respectively. Further, in the resting EEG, when subjects are sitting quietly and not performing a task, alpha-band power over the PFC is relatively higher over the left than the right hemisphere in patients with depression, suggesting a relative hyperactivation of the right PFC. Finally, damage to the left frontal cortex is more likely to produce depression than is a frank lesion of any other part of the brain.

From a neuroscience perspective, LeDoux (2012b) argues that phenomena falling under the rubric of “emotion,” whose neurobiological bases can be studied, include “circuits involved in defense, maintenance of energy and nutritional supplies, fluid balance, thermoregulation, and reproduction. These survival circuits and their adaptive functions are conserved to a significant degree ... across mammalian species, including humans” (p. 654). He argues that it is important that we not conflate such phenomena, however, with humans’ subjective experience of emotions, with “feeling words,” such as *fear, anger, love, sadness, jealousy*, and so on.

By focusing on survival functions instantiated in conserved circuits, key phenomena relevant to emotions and feelings are discussed with the natural direction of brain evolution in mind (by asking to what extent are functions and circuits that are present in other mammals also present in humans) rather than by looking backward, and anthropomorphically, into evolutionary history (by asking whether human emotions/feelings have counterparts in other animals). (p. 654)

THE CONTROL OF EMOTIONS

Extinction

As important as it can be to an organism to learn that a particular stimulus, or context, represents a threat, and should be avoided, it can be equally important to extinguish this association once the threat no longer exists. There is, regrettably, a real-world example of this fact that will be familiar to almost any reader of this book, regardless of where on the planet s/he lives. For a soldier who is in an active combat zone, it is highly adaptive to take evasive action when there is an unexpected loud noise. Often, this is produced by an explosion or by gunfire, and the soldier needs to protect himself and/or do what's necessary to repel the attackers. Once decommissioned and back in civilian life, however, it is no longer adaptive for the ex-soldier to respond in the same way to an unexpected loud sound. The sound is highly likely to be the result of something innocuous, like a car backfiring, or the crash of a dog knocking over a garbage can. “Hitting the dirt” or running for cover in the civilian context,

can interfere with family and other social relations, can make it difficult to obtain or hold a job, and so on. Further along, we'll see that at pathological levels, an inability to extinguish no-longer-relevant conditioned fear can result in post-traumatic stress disorder (PTSD), and possibly other anxiety-related disorders.

Research in the rodent

Fear extinction refers to the gradual decrement in conditioned fear responses that occurs with repeated presentation of a CS without the accompanying US. Extinction was originally thought by Pavlov and his contemporaries to be an inhibitory phenomenon, akin to “unlearning,” or forgetting, the CS–US association. Research from physiological psychology has long-since established, however, that extinction of conditioned fear is better understood as the result of learning a new association with the CS, one that competes with the previously learned CS–US association. Behavioral phenomena such as spontaneous recovery and rapid relearning provide evidence against an “unlearning” mechanism. Consistent with the “new learning” account of extinction is the finding that fear extinction requires the participation of NMDA receptors – i.e., of new learning – in the basal nucleus of the amygdala. The principles and mechanisms that govern fear extinction have now been well worked out in the rodent, and translating these to the human, including into potential treatment strategies, are currently the focus of intensive research activity.

The roles of midline frontal cortex

As indicated above, the basal nucleus of the amygdala has been identified as an important locus for learning a new association for the CS that was previously associated with a threatening US. Learning the new association presumably entails the pairing of sensory information representing the CS with a representation of the context in which extinction trials are administered. (Extinction is more quickly learned if extinction trials – i.e., trials presenting the CS in the absence of the threat-related US – are carried out in a different context from the one in which the fear memory was acquired.) Additionally, however, there are top-down control signals that adjudicate between

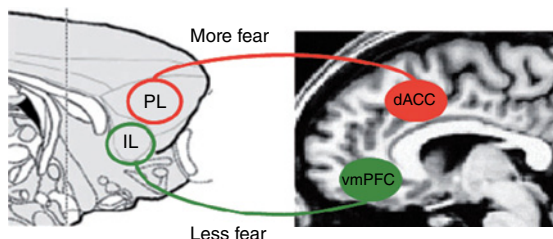


FIGURE 17.7 Anteromedial regions in the rat and the human believed to play homologous roles in the control of conditioned fear memories. (PL = paralimbic; IL = infralimbic). Source: Milad and Quirk, 2012. Reproduced with permission of Annual Reviews.

the older CS–fear association and the newer CS–neutral association. The sources of this top-down control are the midline frontal regions illustrated in *Figure 17.7*. Interestingly, the two seem to operate in antagonistic, “push me–pull you” fashion.

The role for the infralimbic (IL) cortex of the rodent in prioritizing the newer CS–neutral association (often referred to as the “extinction memory”) has been demonstrated in a series of elegant studies by neurobiologist Gregory Quirk and his colleagues at the University of Puerto Rico. The first experiments in the series demonstrated that whereas lesions of IL did not interfere with extinction on day 1 of an experiment, they did interfere with retrieval of this newer CS–neutral association on day 2. That is, the animal’s behavior on day 2 was similar to what it would have been had it not gone through extinction learning on day 1. At face value this could either reflect a deficit in the consolidation of the newly learned CS–neutral association, or an impairment in the newer memory’s retrieval. Evidence in favor of the latter account came from recording studies indicating that whereas IL neurons responded to the CS (a tone) during neither the initial acquisition of the fear memory nor during its extinction, they did respond to the CS during retrieval of the new CS–neutral memory on day 2 of extinction. Furthermore, the magnitude of these “extinction memory” responses was inversely correlated with the amount of freezing at the retrieval test, suggesting that the strength of this IL signal was determining the influence of the newer CS–neutral memory on the animal’s

behavior. Causal evidence for this interpretation was produced by microstimulation of the IL neurons that represented the CS during retrieval of the extinction memory – microstimulation enhanced the expression of the newer CS–neutral memory over the older CS–US fear memory.

As alluded to above, a second region along the frontal midline, the prelimbic (PL) cortex (*Figure 17.7*), seems to play the opposite role. First, inactivation of PL has been shown to reduce the expression of conditioned fear (i.e., the duration of time that the animal freezes upon hearing the tone; *Real-World Applications 10.1*). Second, neurons in PL show sustained increases in firing rate in response to presentation of the CS, and the time course of this activity closely mirrors the duration of the freezing behavior itself. This sustained activity seems to be the product of activity in a recurrent network linking PL with the amygdala, in that the onset of the CS is first detected by the amygdala, which sends a transient “fear signal” to PL. PL then converts this into a sustained signal that feeds back onto the amygdala, resulting in sustained amygdala output to the systems that implement the freezing response (*Figure 17.3*). Finally, as they have done with the IL circuit, Quirk’s group has demonstrated that microstimulation of PL neurons has the effect of increasing freezing responses to the CS and of interfering with extinction.

Fear conditioning and extinction in the human

As suggested previously in the text, and diagrammed explicitly in *Figure 17.7*, there seem to be important parallels from the rodent work that we have just reviewed, and findings with related behaviors in the human. In particular, activity in human vmPFC seems to support functions that are analogous to those of the rodent IL: during fear conditioning, activity in vmPFC is suppressed relative to baseline; during extinction learning, activity in vmPFC increased; and during retrieval of the extinction memory, the magnitude of vmPFC activity predicts the subject’s success at inhibiting conditioned responding. In addition to these functional studies, analysis of structural MRI data suggests that individual differences in the thickness of

Tip 17.3

One (automatic) autonomic response to fear, stress, or anxiety is increased activity in sweat glands in the skin. This is detected as an increase in skin conductance, and is often used as a dependent measure in studies of emotion.

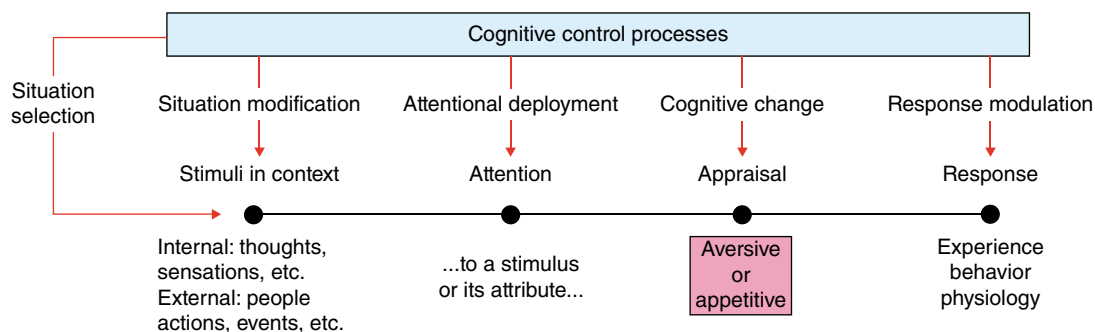
cortex in the vmPFC (as measured with voxel-based morphometry (VBM)) is positively correlated with the strength of extinction learning.

With regard to the establishment and maintenance of conditioned fear memories, there’s good evidence that the human dorsal ACC supports functions that are analogous to those of the rodent PL (*Figure 17.7*): activity in this region correlates with behavioral and skin-conductance (*Tip 17.3*) correlates of fear conditioning, and there’s also a large body of evidence indicating that this region is highly responsive to aversive USs (e.g., the delivery of pain, or an unexpected and loud noise). Interestingly, the omission of an expected US also produces increased activity in this region, an effect that likely reflects a role for error prediction signals in fear conditioning.

“Real-time,” volitional control of emotions

Any particular emotion can influence us for varying lengths of time. Davidson and Begley (2013) summarize it as such: “The smallest, most fleeting unit of emotion is an emotional *state* ... a feeling that ... persist[s] over minutes or hours or even days, is a *mood* ... And a feeling that characterizes you ... for years is an emotional *trait*” (p. xi). In contrast to acquisition and extinction of conditioned fear, processes of which we are typically unaware, and which contribute implicitly to our emotional and personality traits, emotional states and moods are amenable to volitional control. One model of the cognitive control of emotion, which social cognitive neuroscientist Kevin Ochsner and colleagues have been building and refining for over a decade, is illustrated in *Figure 17.8* (*Tip 17.4*). In effect, the model posits that many of the same principles and

A Strategies and processes



B Neural systems

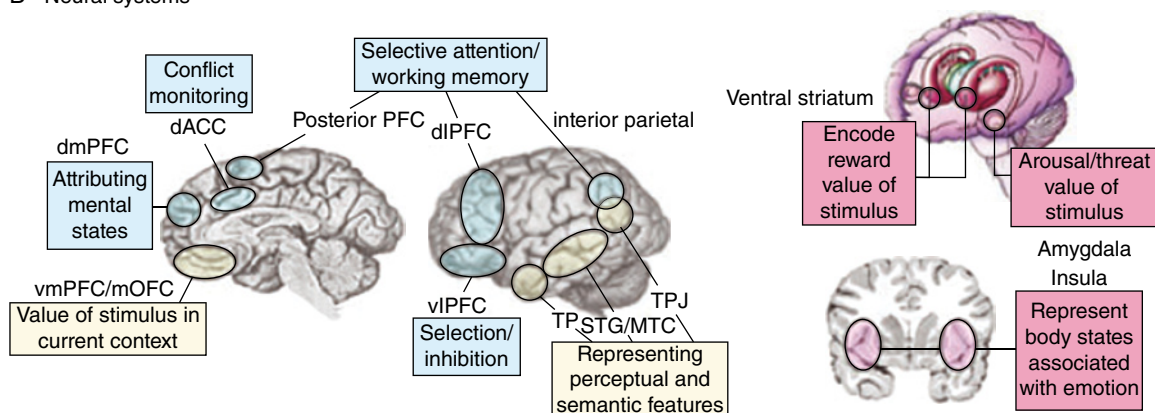


FIGURE 17.8 Ochsner's model of the cognitive control of emotion. In **A**, the four black dots represent the stages involved in generating an emotional state. First, a triggering event occurs, be it a spontaneous thought or an external event. Second, attention is drawn to all or just a part of the representation of the triggering event. Third, a process of appraisal happens, whereby a determination is made about the valence (positive or negative) of this event. There are many theories about the processes that may underlie emotional appraisal, and the neural bases of some processes that may be important appear in blue in panel **B**; here Ochsner, Silvers, and Buhle (2012) have simplified the appraisal step into a binary classification of “aversive or appetitive.” (Note that this has parallels with Davidson’s “avoid/approach” distinction.) Finally, the “Response” stage corresponds to “translating these appraisals into changes in experience, emotion-expressive behavior, and autonomic physiology” (p. e3). The control processes that can influence each of these stages are elaborated in the text. Panel **B** summarizes mental processes and neural systems proposed to be involved in appraisal (blue), generating responses (pink), and systems with “an undefined or intermediary role in reappraisal” (p. e3; yellow). Source: Ochsner, Silvers, and Buhle, 2012. Reproduced with permission of New York Academy of Sciences.

mechanisms that we use for cognitive control are also used for emotion regulation (*Tip 17.5*).

In the model of emotion regulation illustrated in *Figure 17.8*, individual mechanisms that implement

emotion-regulation each act on one of the steps in emotion generation that are illustrated in panel **A** of *Figure 17.8*. “Situation selection” and “situation modification” are strategies for avoiding an encounter with a

Tip 17.4

There's a nice one-page summary of "What is Social Cognitive Neuroscience?," including a philosophy, summary of principal journals and conferences, etc., on the home page of Ochsner's laboratory's Web page (Social Cognitive Neuroscience Laboratory, n.d.).

Tip 17.5

Note that the convention in the field is to refer to emotion *regulation* and to cognitive *control*, even though the two words really relate to the same class of operation.

potentially aversive situation. "Attentional selection" and its complement, "attentional filtering," control what information is gated into, or out of, the emotion generation process. "Cognitive change" refers to a collection of processes and strategies that act on the process of emotional appraisal. The most widely studied of this is "cognitive reappraisal," which entails reinterpreting the meaning or significance of the triggering event, so as to change one's emotional response to it. Finally, the processes that act on the expression of emotion ("Response" in *Figure 17.8*) fall under the rubric of "response modulation." The most commonly studied of these is expressive suppression – controlling the muscles of one's face to disguise from observers what emotion one is experiencing (e.g., the "poker face").

The good and the bad of emotion regulation

Our emotional states, moods, traits, are arguably the most important factors in determining our quality of life. The "good" in this section title refers to the growing body of promising findings indicating that through intensive, scientifically informed training, one can become more skilled at regulating one's emotional states, and reap collateral cognitive benefits,

to boot. Some evidence for this is summarized in *Methodology Box 17.1*. The "bad" refers to the fact that affective neuroscience research has shown conclusively that many types of psychopathology – depression, PTSD, generalized anxiety disorder, to name just a few – can be linked to dysfunction of emotion regulation. But, to engage in some quick reappraisal, understanding how dysfunctional emotion regulation leads to "bad" mental health is a positive development, because, of course, understanding the factors underlying a disorder is a necessary first step to developing effective treatments. The role of emotion regulation in psychopathology is considered in *Real-World Applications 17.2*.

THE AMYGDALA AS AFFECTIVE HUB

At the beginning of this chapter we considered two topics that may not be traditionally thought of as "emotional" – trustworthiness and racial stereotyping. However, when considered from the perspective of Davidson's approach/avoidance framework, these might both be classified as avoidance-related impulses. Additionally, because they are learned, and can be "unlearned" (Phelps et al., 2000; *Real-World Applications Box 17.1*), and because of how they engage the amygdala, it seems that they follow the same rules as do the conditioning and extinction of unambiguously fearful memories.

More generally, we have seen that the amygdala engages with other brain systems in ways that are reminiscent of the interactivity that we emphasized in relation to visual, auditory, and somatosensory perception. In one example, we have seen that to sustain a conditioned fear memory, the amygdala engages a recurrent circuit with the dorsal ACC. In a second, we have seen several examples of how detection of a fast, coarse signal from a feedforward visual channel triggers the amygdala to send feedback signals that boost activity in visual cortex. This latter scenario is an example of the top-down signals that we invoked in *Chapter 3* to account for the dense feedback control that extrastriate areas impose on V1, and that V1, in turn, imposes on the LGN.

METHODOLOGY BOX

17.1 Meditation as mental training

With an introduction seemingly more fitting for this book's *Chapter 8* than *17*, here's how cognitive neuroscientist Heleen Slagter and colleagues (2007; including senior author Richard Davidson) set up their meditation study:

A major limitation in human information processing arises from the time required to consciously identify and consolidate a visual stimulus in short-term memory ... This process can take more than a half-second before it is free for a second stimulus, as is revealed by the attentional-blink paradigm: If the second target stimulus (T2) of two target stimuli is presented within 500 ms of the first one (T1) in a rapid sequence of distracters, it is often not detected ... This deficit ... is believed to result from competition between the two targets for limited attentional resources ...: When many attentional resources are devoted to the processing of T1, too few may be available for T2 ... Yet, the attentional blink does not reflect a general, immutable bottleneck, because most individuals are able to identify T2 on at least a portion of trials ... suggest[ing] that some control ... over the allocation of attentional resources is possible. (p. 1)

Subjects performed the attentional-blink task, with concurrent EEG, immediately prior to beginning, and immediately upon completing, a three-month retreat during which they meditated for 10–12 hours per day with Vipassana meditation. With this technique one starts by focusing or stabilizing concentration on an object such as the breath, and subsequently broadens one's focus, cultivating a “non-reactive” form of sensory awareness, meaning that one does not respond emotionally to the foci of attention. Control subjects were individuals who were interested in learning about meditation; they were given a one-hour class, and meditated for 20 minutes daily for the week prior to each EEG session.

Behavioral results showed a greater reduction of the attentional blink post- vs. pre-training for the meditation group relative to the control group. Notably, however, both groups improved slightly at T1 detection on session 2 vs. session 1, and the control group also showed a modest (though significantly smaller) improvement at T2 detection. This highlights the importance of including a control group in training studies. Without one, one cannot rule out effects of learning the task, or some other factor that may have nothing to do with the training procedure itself. The ERP data, illustrated in *Figure 17.9*, also showed a selective effect of meditation training.

Although other studies have shown that intensive performance of a task can produce similar reductions in the attentional blink (e.g., *Web Link 15.1*), noteworthy about the present finding is that the benefits of training resulted from *purely mental* activity – subjects weren't pointing a gun, controlling a joystick, or performing any other overtly sensorimotor task; they were “just thinking.” (“Just thinking,” of course, in a particularly disciplined, focused way.) (For more detail, see *Web Clip 17.2*.) Subsequent studies from Davidson and colleagues have demonstrated impressive effects with other types of mental training. One example is that two weeks of compassion training (relative to a reappraisal control group) alters performance on a neuroeconomic game entailing the altruistic redistribution of funds, a behavioral change accompanied by changes in task-related functional connectivity between dorsolateral PFC and the nucleus accumbens (Weng et al., 2013).

Methodology Box 17.1 *Continued*

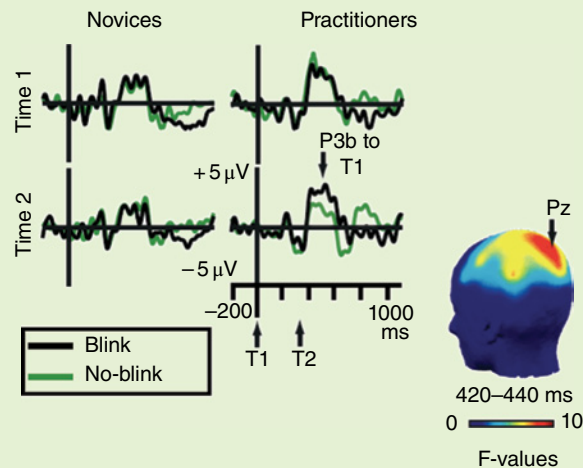


FIGURE 17.9 Comparison of the P3b ERP component evoked by the T1 stimulus as a function of group (Novices (i.e., controls); Practitioners (i.e., meditation training group)); pre- vs. post-training session (“Time 1” and “Time 2,” respectively) and, in each plot, whether the following T2 stimulus was detected (“No-blink”) or missed (“Blink”). The P3b is associated with the allocation of attentional resources. The finding of a lower P3b to “No-blink” T1 stimuli in the post-training session for Practitioners, but not Novices, therefore, is interpreted as evidence that, as a result of training, they were able to devote fewer attentional resources to T1, thereby leaving more available for detection of T2. Also consistent with this interpretation was the fact that the individuals who showed the largest Time 1-to-Time 2 reductions in P3b were those who showed the largest reductions in attentional blink from Time 1 – Time 2. Source: Slagter, Lutz, Greischar, Francis, Nieuwenhuis, Davis, and Davidson, 2007. Reproduced with permission of the authors.

REAL-WORLD APPLICATIONS

17.2 Emotion regulation gone awry: causes, and possible treatments, of anxiety disorders

In this chapter we have considered two classes of emotion regulation: the extinction of Pavlovian conditioning (with an emphasis on fear conditioning) and the volitional control of emotional states. Dysfunction of both mechanisms is implicated in various mental illnesses.

Post-traumatic stress disorder (PTSD), as defined by the (US) National Institute of Mental Health, “develops after a terrifying ordeal that involved physical harm or the threat of physical harm.” It is associated with three classes of symptoms: *re-experiencing* (e.g., flashbacks, nightmares); *avoidance* (staying away from places, events, or objects that are reminders of the traumatic experience); and *hyperarousal* (chronically elevated anxiety, easily startled, angry outbursts). There is now considerable evidence linking PTSD to

Real-World Applications 17.2 Continued

fear conditioning that is resistant to extinction. For example, research by former Quirk student Mohammed Milad and colleagues (2009) indicates that PTSD patients show normal extinction on the day that extinction trials are administered, but a deficit in extinction recall on the day after extinction trials. Concurrent fMRI shows this return of the conditioned association to be accompanied by hypoactivation in the vmPFC and hyperactivation in the dACC (see more at *Web Link 17.2*). (Recall that an analogous pattern was observed in rats with vmPFC lesions; *Figure 17.7*.) One potentially promising form of treatment for PTSD is also derived from research in the rodent: just as the activation of NMDA receptors in the amygdala is necessary for fear extinction, preliminary reports from the clinic have suggested that the efficacy of cognitive behavioral therapy as a treatment for PTSD is enhanced by the concurrent administration of an NMDA agonist.

Turning to the volitional control of emotional state and mood, there is now a large body of evidence associating depression and other anxiety-related disorders with abnormalities in one or more of the emotion regulation processes summarized in *Figure 17.8*. For example, anhedonia, the loss of pleasure or interest in previously rewarding stimuli or behaviors, is a cardinal symptom of **major depressive disorder (MDD)**. Although many classical theories of depression posit that the disease is characterized by a reduced hedonic capacity – in effect, a limit to the amount of pleasure one can experience – others have emphasized impaired control, such as the ability to sustain positive affect and reward responsiveness over time. In one study by Heller and colleagues (2009), subjects were instructed to use cognitive (re)appraisal to “enhance” or “suppress” their emotional response to positive and negative images. It was found that, in patients with MDD, activity in the nucleus accumbens decreased over the time while they were to “enhance” positive emotional responses, whereas in control subjects activity levels remained elevated. In **generalized anxiety disorder (GAD)**, worry is a cardinal symptom. An fMRI study by Nitschke et al. (2009) indicated that patients with GAD showed abnormally high levels of activity in the amygdala during the anticipation period before a negative image was presented. Following the scan, patients began taking the anti-anxiety medication venlafaxine. Interestingly, correlations of treatment response to neural activity implicated a different region: patients who had higher levels of anticipatory activity in ACC evinced greater drug-related reductions in anxiety and worry symptoms. This suggests a potential way to determine ahead of time whether a patient is likely to respond to a particular pharmacological intervention.

END-OF-CHAPTER QUESTIONS

1. In what way(s) does it make sense to think of faces and the information that they convey – about emotion, trustworthiness, race – as conditioned stimuli?
2. Trace the circuitry that underlies fear conditioning. Is this the same or different from the circuitry underlying avoidance learning?
3. Fear conditioning and avoidance learning are both types of nondeclarative memory. Is the amygdala also involved in processing the emotional component of declarative memories? If so, through what mechanism(s)?
4. How, and under what circumstances, does the amygdala influence visual processing?
5. What is an emotion? How are emotions different, neurobiologically, from the domains of cognition that we have studied up through this point in the book?
6. Is the role of the vmPFC in fear extinction best understood as learning the new CS, or retrieving it after it is learned? Contrast this with the role of the ACC.

7. Think back to the finding, summarized in *Real-World Application 16.1* and *Figure 16.3*, relating the structure of the uncinate fasciculus to psychopathy. What predictions might you make about variation in this structure as a function of individual differences in the ability to extinguish conditioned fear memories?
8. How do the processes of emotion regulation differ from extinction? What brain systems and mechanisms are associated with emotion regulation?
9. How is emotion regulation related to psychopathology?

REFERENCES

- Adolphs, Ralph, Larry Cahill, Rina Schul, and Ralf Babinsky. 1997. "Impaired Declarative Memory for Emotional Material following Bilateral Amygdala Damage in Humans." *Learning & Memory* 4 (3): 291–300. doi: 10.1101/lm.4.3.291.
- Adolphs, Ralph, Daniel Tranel, and Antonio R. Damasio. 1998. "The Human Amygdala in Social Judgment." *Nature* 393: 470–474. doi: 10.1038/30982.
- Cahill, Larry, Richard J. Haier, James Fallon, Michael T. Alkire, Cheuk Tang, David Keator, ... James L. McGaugh. 1996. "Amygdala Activity at Encoding Correlated with Long-Term, Free Recall of Emotional Information." *Proceedings of the National Academy of Sciences of the United States of America* 93 (15): 8016–8021. doi: 10.1073/pnas.93.15.8016.
- Cahill, Larry, Bruce Prins, Michael Weber, and James L. McGaugh. 1994. "β-Adrenergic Activation and Memory for Emotional Events." *Nature* 371 (6499): 702–704. doi: 10.1038/371702a0.
- Davidson, Richard J., and Sharon Begley. 2013. *The Emotional Life of Your Brain: How Its Unique Patterns Affect the Way You Think, Feel, and Live – and How You Can Change Them*. New York: Plume.
- Davidson, Richard J., and William Irwin. 1999. "The Functional Neuroanatomy of Emotion and Affective Style." *Trends in Cognitive Sciences* 3 (1): 11–21. doi: 10.1016/S1364-6613(98)01265-0.
- Engell, Andrew D., James V. Haxby, and Alexander Todorov. 2007. "Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala." *Journal of Cognitive Neuroscience* 19 (9): 1508–1519.
- Hariri, Ahmad R., and Paul J. Whalen. 2011. "The Amygdala: Inside and out." *F1000 Biology Reports* 3: 2. doi: 10.3410/B3-2.
- Heller, Aaron S., Tom Johnstone, Alexander J. Shackman, Sharee N. Light, Michael J. Peterson, Gregory G. Kolden, Ned H. Kalin, and Richard J. Davidson. 2009. "Reduced Capacity to Sustain Positive Emotion in Major Depression Reflects Diminished Maintenance of Fronto–Striatal Brain Activation." *Proceedings of the National Academy of Sciences of the United States of America* 106 (52): 22445–22450. doi: 10.1073/pnas.0910651106.
- Klüver, Heinrich. 1926. "Mescal Visions and Eidetic Vision." *American Journal of Psychology* 37 (4): 502–515.
- Klüver, Heinrich. 1928. *Mescal: The "Divine" Plant and Its Psychological Effects*. London: Kegan Paul, Trench, Trubner.
- LeDoux, Joseph E. 2012a. "Evolution of Human Emotion: A View through Fear." In *Evolution of the Primate Brain: From Neuron to Behavior*, edited by Michael A. Hofman and Dean Falk, vol. 195 of Progress in Brain Research, edited by Stephen G. Waxman, Donald G. Stein, Dick F. Swaab, and Howard L. Fields, 431–442. Amsterdam: Elsevier. doi: 10.1016/B978-0-444-53860-4.00021-0.
- LeDoux, Joseph E. 2012b. "Rethinking the Emotional Brain." *Neuron* 73 (4): 653–676. doi: 10.1016/j.neuron.2012.02.004.
- Milad, Mohammed R., Roger K. Pitman, Cameron B. Ellis, Andrea L. Gold, Lisa M. Shin, Natasha B. Lasko, ... Scott L. Rauch. 2009. "Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder." *Biological Psychiatry* 66 (12): 1075–1082. doi: 10.1016/j.biopsych.2009.06.026.
- Milad, Mohammed R., and Gregory J. Quirk. 2012. "Fear Extinction As a Model for Translational Neuroscience: Ten Years of Progress." *Annual Review of Psychology* 63: 129–151. doi: 10.1146/annurev.psych.121208.131631.
- Nahm, Frederick K. D., and Karl H. Pribram. 1998. "Heinrich Klüver 1897–1979." In *Biographical Memoirs* 73: 1–19. Washington, DC: National Academies Press.
- Nitschke, Jack B., Issidoros Sarinopoulos, Desmond J. Oathes, Tom Johnstone, Paul J. Whalen, Richard J. Davidson, and Ned H. Kalin. 2009. "Anticipatory Activation in the Amygdala and Anterior Cingulate in Generalized Anxiety Disorder and Prediction of

- Treatment Response.” *American Journal of Psychiatry* 166 (3): 302–310. doi: 10.1176/appi.ajp.2008.07101682.
- Ochsner, Kevin N., Jennifer A. Silvers, and Jason T. Buhle. 2012. “Functional Imaging Studies of Emotion Regulation: A Synthetic Review and Evolving Model of the Cognitive Control of Emotion.” *Annals of the New York Academy of Sciences* 1251: e1–e24. doi: 10.1111/j.1749-6632.2012.06751.x.
- Phelps, Elizabeth A., Kevin J. O’Connor, William A. Cunningham, E. Sumie Funayama, J. Christopher Gatenby, John C. Gore, and Mahzarin R. Banaji. 2000. “Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation.” *Journal of Cognitive Neuroscience* 12 (5): 729–738. doi: 10.1162/089892900562552.
- Slagter, Heleen A., Antoine Lutz, Lawrence L. Greischar, Andrew D. Francis, Sander Nieuwenhuis, James M. Davis, and Richard J. Davidson. 2007. “Mental Training Affects Distribution of Limited Brain Resources.” *PLoS Biology* 5 (6): e138. doi: 10.1371/journal.pbio.0050138.
- Social Cognitive Neuroscience Laboratory. N.d. “What is Social Cognitive Neuroscience?” <http://dept.psych.columbia.edu/~kochsner/home.htm>.
- Vuilleumier, Patrik, Mark P. Richardson, Jorge L. Armony, Jon Driver, and Raymond J. Dolan. 2004. “Distant Influences of Amygdala Lesion on Visual Cortical Activation during Emotional Face Processing.” *Nature Neuroscience* 7 (11): 1271–1278. doi: 10.1038/nn1341.
- Weng, Helen Y., Andrew S. Fox, Alexander J. Shackman, Diane E. Stodola, Jessica Z. K. Caldwell, Matthew C. Olson, ... Richard J. Davidson. 2013. “Compassion Training Alters Altruism and Neural Responses to Suffering.” *Psychological Science* 24 (7): 1171–1180. doi: 10.1177/0956797612469537.
- Whalen, Paul J., Jerome Kagan, Robert G. Cook, F. Caroline Davis, Hackjin Kim, Sara Polis, ... Tom Johnstone. 2004. “Human Amygdala Responsivity to Masked Fearful Eye Whites.” *Science* 306 (5704): 2061. doi: 10.1126/science.1103617.

OTHER SOURCES USED

- Adolphs, Ralph, Daniel Tranel, Hanna Damasio, and Antonio R. Damasio. 1994. “Impaired Recognition of Emotion in Facial Expressions following Bilateral Damage to the Human Amygdala.” *Nature* 372 (6507): 669–672. doi: 10.1038/372669a0.
- Cunningham, William A., Marcia K. Johnson, Carol L. Raye, J. Chris Gatenby, John C. Gore, and Mahzarin R. Banaji. 2004. “Separable Neural Components in the Processing of Black and White Faces.” *Psychological Science* 15 (12): 806–813. doi: 10.1111/j.0956-7976.2004.00760.x.
- Davidson, Richard J. 1998. “Affective Style and Affective Disorders: Perspectives from Affective Neuroscience.” *Cognition and Emotion* 12 (3): 307–330. doi: 10.1080/026999398379628.
- Demos, Kathryn E., William M. Kelley, S. L. Ryan, and Paul J. Whalen. 2008. “Human Amygdala Sensitivity to the Pupil Size of Others.” *Cerebral Cortex* 18 (12): 2729–2734. doi: 10.1093/cercor/bhn034.
- Johnstone, Tom, Carien M. van Reekum, Heather L. Urry, Ned H. Kalin, and Richard J. Davidson. 2007. “Failure to Regulate: Counterproductive Recruitment of Top-Down Prefrontal-Subcortical Circuitry in Major Depression.” *Journal of Neuroscience* 27 (33): 8877–8884. doi: 10.1523/JNEUROSCI.2063-07.2007.
- Klüver, Heinrich, and Paul C. Bucy. 1937. “‘Psychic Blindness’ and Other Symptoms following Bilateral Temporal Lobectomy in Rhesus Monkeys.” *American Journal of Physiology* 119 (2): 352–353.
- LeDoux, Joseph E. 2013. “The Slippery Slope of Fear.” *Trends in Cognitive Sciences* 17 (4): 155–156. doi: 10.1016/j.tics.2013.02.004.
- Phan, K. Luan, Mike Angstadt, Jamie Golden, Ikechukwu Onyewuenyi, Ana Popovska, and Harriet de Wit. 2008. “Cannabinoid Modulation of Amygdala Reactivity to Social Signals of Threat in Humans.” *Journal of Neuroscience* 28 (10): 2313–2319. doi: 10.1523/JNEUROSCI.5603-07.2008.
- Reznikoff, Glen A., Scott Manaker, C. Harker Rhodes, Andrew Winokur, and Thomas C. Rainbow. 1986. “Localization and Quantification of Beta-Adrenergic Receptors in Human Brain.” *Neurology* 36 (8): 1067–1073. doi: 10.1212/WNL.36.8.1067.
- Vuilleumier, Patrik, Jorge L. Armony, Jon Driver, and Raymond J. Dolan. 2001. “Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-Related fMRI Study.” *Neuron* 30 (3): 829–841. doi: 10.1016/S0896-6273(01)00328-2.
- Vuilleumier, Patrik. 2005. “How Brains Beware: Neural Mechanisms of Emotional Attention.” *Trends in Cognitive Sciences* 9 (12): 585–594. doi: 10.1016/j.tics.2005.10.011.

FURTHER READING

- Bavelier, Daphne, and Richard J. Davidson. 2013. "Brain Training: Games to Do You Good." *Nature* 494 (7438): 425–426. doi: 10.1038/494425a.
Leading experts in the training of attention and emotion, respectively, consider the question of "Because gaming is clearly here to stay, ... how to channel people's love of screen time towards positive effects on the brain and behavior."
- Davidson, Richard J., and Sharon Begley. 2013. *The Emotional Life of Your Brain: How Its Unique Patterns Affect the Way You Think, Feel, and Live – and How You Can Change Them*. New York: Plume.
An engrossing, authoritative "users' manual" for understanding and controlling one's emotional style, interwoven with facts from affective neuroscience and anecdotes from the career of one of its leading scholars.
- Davis, Michael, and Paul J. Whalen. 2001. "The Amygdala: Vigilance and Emotion." *Molecular Psychiatry* 6 (1): 13–34.
Detailed review of the anatomy, physiology, and pharmacology of conditioning in the amygdala.
- Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102 (1): 4–27. doi: 10.1037/0033-295X.102.1.4.
This paper lays out the theoretical background for understanding implicit bias and stereotyping from a social psychology perspective.
- Kubota, Jennifer T., Mahzarin R. Banaji, and Elizabeth A. Phelps. 2012. "The Neuroscience of Race." *Nature Neuroscience* 15 (7): 940–948. doi: 10.1038/nn.3136.
A review of exciting, vitally important research that has revealed "overlap in the neural circuitry of race, emotion and decision-making."
- LeDoux, Joseph E. 2000. "Emotion Circuits in the Brain." *Annual Review of Neuroscience* 23: 155–184.
Provocative argument by one of the leading scientists of fear conditioning for reconceptualizing emotion research as research on survival circuits.
- Milad, Mohammed R., and Gregory J. Quirk. 2012. "Fear Extinction As a Model for Translational Neuroscience: Ten Years of Progress." *Annual Review of Psychology* 63: 129–151. doi: 10.1146/annurev.psych.121208.131631.
A review of the truly pioneering research, much of it performed by the authors, on the neural bases of extinction learning.
- Ochsner, Kevin N., and Matthew D. Lieberman. 2007. "The Emergence of Social Cognitive Neuroscience." *American Psychologist* 56 (9): 717–734. doi: 10.1037//0003-066X.56.9.717.
As told by two of the field's leading practitioners.

CHAPTER 18

LANGUAGE

KEY THEMES

- Language, with its ubiquity in our lives, and its uniqueness to our species, is the focus of strongly debated theories that get at the core of what it means to be human.
- Nineteenth-century neurological models posited a network comprising posterior temporal cortex (Wernicke's area), temporoparietal regions, posterior inferior frontal gyrus (Broca's area), and, critically, the connections between them, that still influences contemporary ideas of the neural organization of language.
- EEG indicates that the earliest cortical generators of the AEP begin segregating speech from nonspeech sounds, probably bilaterally.
- The idea that there may exist a single "grammar gene" has prompted considerable excitement and research, but is probably "too good to be true."
- More than 150 years after Broca's first report of language impairment following damage to inferior PFC, the function(s) attributable to "his" area are still the focus of contentious debate.
- EEG and MEG suggest that predictive coding is an important principle governing language processing; both its comprehension and its production.
- The study of dyslexia, a developmental disorder of reading, reveals critical roles for phonological processing and "phonological awareness," even for this overtly visual aspect of language.

CONTENTS

KEY THEMES

A SYSTEM OF REMARKABLE COMPLEXITY

WERNICKE–LICHTHEIM: THE CLASSICAL CORE LANGUAGE NETWORK

The aphasias

The functional relevance of the connectivity of the
network

SPEECH PERCEPTION

Segregation of the speech signal

The elements of language

The speech signal

Neural correlates of speech perception

Dual routes for speech processing

GRAMMAR

Genetics

Rules in the brain?

Broca's area

The anatomy and connectivity of Broca's area

Functional accounts of Broca's area

The electrophysiology of grammar

SPEECH PRODUCTION

A psycholinguistic model of production

Forward models for the control of production

Necessary neural substrates?

ALL TOGETHER NOW

“AROUSING NOVEL IDEAS AND SUBTLE PERCEPTIONS AND JUDGMENTS” ... WITH THE WRITTEN WORD

END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

A SYSTEM OF REMARKABLE COMPLEXITY

A human language is a system of remarkable complexity. To come to know a human language would be an extraordinary intellectual achievement for a creature not specifically designed to accomplish this task. A normal child acquires this knowledge on relatively slight exposure and without specific training. He can then quite effortlessly make use of an intricate structure of specific rules and guiding principles to convey his thoughts and feelings to others, arousing in them novel ideas and subtle perceptions and judgments. (Chomsky, 1975, p. 4)

Language is, indeed, a most remarkable element of human cognition. It is the medium through which culture is transmitted, technology propagated, and knowledge about topics as rich and complex as, say, *cognitive neuroscience*, is taught, updated, debated, refined, and so on. It also seems to be a uniquely human faculty. Although some other species have sophisticated means of communication, and some can be taught some rudimentary elements of human language, none has a system of communication whereby a finite number of elements can be recombined in an infinite number of ways, to generate an infinite number of messages. This is referred to as the generative property of human language.

Perhaps it is because language is so uniquely human, and so intertwined with virtually everything that we do as individuals, and as societies, that it is the most intensively studied domain of human cognition (*Tip 18.1*). One result of this intense interest is that there are many

Tip 18.1

Of the 17 domains of human cognition and behavior that each get a chapter in this book (i.e., *Chapters 3–19*), *Language* is the only one to which entire university departments are dedicated (e.g., Linguistics; Speech and Communicative Disorders). Additionally, language is also commonly represented as a subdiscipline of such scholarly fields as philosophy, evolutionary biology, and anthropology.

theoretical debates about, for example, how language is learned, the rules of grammar (indeed, whether there even exist rules of grammar), and the extent to which language depends on neural structures that are unique to our species. Although these debates can be rich and fascinating, and often get at core questions about human nature, many of them are simply outside the remit of this book. Nonetheless, when appropriate, theoretical motivations and/or implications of particular studies that we will consider will be flagged.

WERNICKE–LICHTHEIM: THE CLASSICAL CORE LANGUAGE NETWORK

The aphasias

In *Chapter 1*, we introduced the work of Broca, including with his famous patient “Tan,” as an important demonstration, alongside roughly contemporaneous research on motor control and vision, of localization of function in the brain. In his famous and influential paper from 1970, the neurologist Norman Geschwind (1926–1984) pointed out another important feature of this work: “The aphasias were ... the first demonstrations of the fact that selective damage to the brain could affect one class of *learned behavior* while sparing other classes” (p. 940, emphasis added). That is, the first evidence that an unequivocally cognitive function (as contrasted with motor control and vision) could be localized to a particular part of the brain. But “the aphasias”? Broca only described one. What Geschwind was also referring to was the work of the Germany-based neurologist Carl Wernicke (1848–1904) who, in 1874, proposed a model for the neural bases of two classes of aphasia. Here’s how Geschwind (1970) summarized them:

The aphasic of the Broca’s type characteristically produces little speech, which is emitted slowly, with great effort, and with poor articulation ... Characteristically the small grammatical words and endings are omitted ... These patients invariably show a comparable disorder in their written output, but they may comprehend spoken and written language normally. In striking contrast to these performances, the patient may retain his musical capacities. It is a common but most dramatic finding to

observe a patient who produces single substantive words with great effort and poor articulation and yet sings a melody correctly and even elegantly.

(For more on the differences, and similarities, of the neural processing of language vs. music, visit *Web Link 18.1*.)

This contrasts with Geschwind's summary of the aphasia first described by Wernicke in 1874:

The Wernicke's aphasi[a] ... patient usually has no paralysis of the opposite side, a fact which reflects the difference in the anatomical localization of his lesion. The speech output can be rapid and effortless, and in many cases the rate of production of words exceeds the normal. The output has the rhythm and melody of normal speech, but it is remarkably empty and conveys little or no information. The patient uses many filler words, and the speech is filled with circumlocutions. There may be many errors in word usage, which are called paraphasias ... The Wernicke's aphasic may, in writing, produce well-formed letters, but the output exhibits the same linguistic defects which are observed in the patient's speech. He shows a profound failure to understand both spoken and written language, although he suffers from no elementary impairment of hearing or sight. (pp. 940–941)

The loci of the regions described by Broca and by Wernicke, and redescribed by Geschwind (1970), are illustrated in *Figure 18.1*. Broca's area is located at the

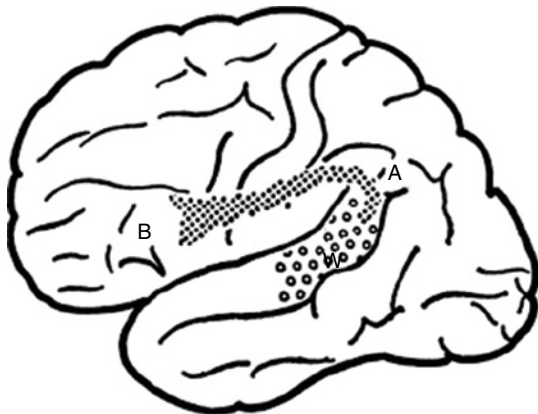


FIGURE 18.1 Geschwind's (1970) depiction of the Wernicke model of the neural bases of language function. (W=Wernicke's area; A=arcuate fasciculus; B=Broca's area.) Source: Geschwind, 1970. Reproduced with permission of AAAS.

posterior end of the inferior frontal gyrus of the left hemisphere of the PFC, immediately anterior to premotor cortical regions that control the face, tongue, lips, palate, and vocal cords. Classically, because of the characteristics of Broca's aphasia, Broca's area has been presumed to be the neural seat of language production (but see *Further Reading* by Dronkers, 2000). Wernicke's area is located in the posterior superior temporal gyrus (STG) in the left hemisphere, occupying a portion of auditory parabelt cortex (*Figure 4.7*). It is, therefore, anatomically proximal to auditory cortex, and receives projections from A1 and belt cortex, as well as direct projections from the medial geniculate nucleus (MGN). Its function has classically been associated with the comprehension of spoken language. (Videos of patients with these classic aphasia are viewable from *Web Link 18.2* and *Web Link 18.3*)

The functional relevance of the connectivity of the network

Although Geschwind didn't include him in the 1970 review, a third important nineteenth-century contributor to our understanding of the brain bases of language function was German neurologist Ludwig Lichtheim (1845–1928). Lichtheim is famous for his “house model,” which diagrammatically predicted how interruptions of connections between input, semantic, and output regions would produce aphasias that would differ in subtle ways from those resulting from frank damage that was limited to Wernicke's or Broca's area (*Figure 18.2*). Remarkably, over the next century, each of the aphasias predicted by Wernicke and by Lichtheim came to be identified by subsequent generations of neurologists. To name just one example, conduction aphasia presents as fluent but **paraphasic speech**, but in the presence of intact language comprehension. (Wernicke's aphasia, in contrast, would present as fluent but paraphasic speech, but in the presence of severely impaired language comprehension.) It is most often associated with damage to the inferior parietal cortex – near the “A” in *Figure 18.1*; and corresponding to “3” in *Figure 18.2.A* – and is explained by damage to the arcuate fasciculus, which has the effect of leaving both Wernicke's and Broca's areas intact, but interrupting the fibers that connect them (*Tip 18.2*).

FIGURE 18.2 Lichtheim's (1885) "house model" of the neural substrates of language. Source: Compston, 2006. Reproduced with permission of Oxford University Press.

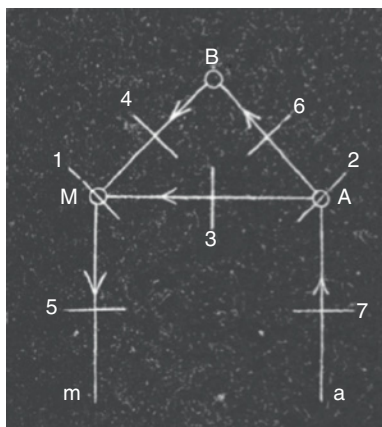


FIGURE 18.2.A "a" and "m" correspond to auditory inputs and motor outputs, respectively; "A" to Wernicke's area and "M" to Broca's. "B" is a stand-in for the neural representation of the semantic content carried by language (see Figure 18.2.B for more elaboration on "B"). The numerals indicate lesioning points, each known (1, 2) or (accurately) predicted (3–7) to produce a different aphasia: 1 = Broca's; 2 = Wernicke's; 3 = conduction; 4 = transcortical motor; 5 = subcortical motor; 6 = transcortical sensory; 7 = subcortical sensory (Tip 18.2). Detail on these can be found in the *Further Reading* by Alexander (1997).

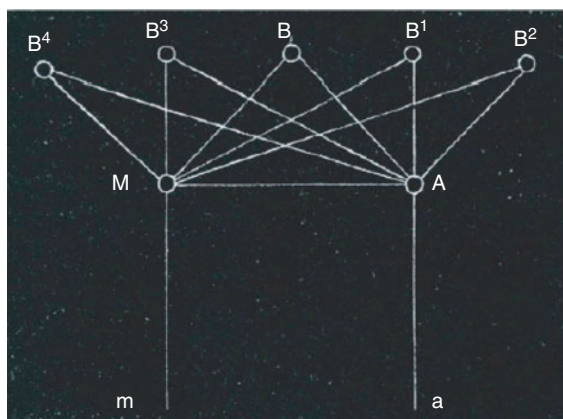


FIGURE 18.2.B Lichtheim's strikingly modern conceptualization of the distributed nature of the brain's representation of knowledge and its anatomical interfaces with Broca's and Wernicke's areas. (Compare, for example, with *Methodology Box 9.1* and *Figure 12.1*.)

Tip 18.2

From the perspective of the Wernicke–Lichtheim model, it's easy to see why a behavioral neurologist might give the label "primary progressive aphasia" (see *Chapter 12*) to a syndrome resulting from degenerative disease affecting "B" in Figure 18.2.

Thus, this Wernicke–Lichtheim model provides a framework from which we can start examining language functions more closely, beginning with speech perception, then proceeding to grammar and production.

SPEECH PERCEPTION

Segregation of the speech signal

In *Chapter 4*, we considered in some detail the processes of the transduction of acoustic energy into a neural code, and the stages of processing of the auditory signal from brainstem to MGN to cortex that are discernible in the auditory evoked potential (AEP). This gives us a starting point to think about speech perception and, in particular, where and how auditory signal that is related to speech is identified and extracted from the auditory stream for its presumably specialized processing as language. A straightforward way to get at this question would clearly be to contrast the processing of speech sounds vs. nonspeech sounds while measuring sound-evoked activity, and determining how the two physiological signals differ. Before just jumping in, however, there are some basic facts that we need to understand about the signal that we want to study. First, what are the elements of spoken language? Second, at what level of organization do we want to study it? Finally, what are some fundamental physical and psychophysical properties of the speech signal itself?

The elements of language

One way to think about the organization of spoken language is as a hierarchy. At the lowest level are **phonemes**, the smallest unit of sound in a language

that, when changed, can change the meaning of an utterance. For example, as we shall see in the next section, in US English the words “doll” and “tall” only differ by one phoneme, even though they differ by two letters. Note that, with a few exceptions, phonemes do not, by themselves, convey meaning. Rather, phonemes are assembled to create words. All human languages use many fewer phonemes than there are possible sounds that can be made with the articulatory apparatus (i.e., the diaphragm, vocal cords, tongue, jaw, and lips). Although the number of phonemes differs across languages, and often differs across dialects within a language, no natural language is believed to contain more than 150. The next level in the hierarchy is *words*. Words are the smallest element of the language that have inherent meaning. The mental representation of words is referred to as the mental lexicon, and the mental lexicon of the typical adult comprises anywhere between 50,000 and 75,000 words. The meaning of a word can be changed by the addition of one or more *morphemes*: prefixes and suffixes that modify words. For example, in English, adding an ‘s’ to the end of a regular noun makes it plural, and adding “un” to the front of one adds negation or “the opposite of.”

The next level is the *clause*. A clause is constructed of words, and it is at this level of language that grammar comes into play. For example, it is the rules of grammar that determine that different ordering of the same three words can have different meanings (e.g., “Dog bites man” vs. “Man bites dog”), or no meaning (e.g., “Bites dog man”). Within the two grammatical clauses just given, the meaning was changed by changing which noun filled the role of *noun phrase* and which the role of *verb phrase*. These two clauses are also each a *sentence*.

Sentences can consist of single clauses or of the combination of two or more clauses. Indeed, it was to the infinite “combine-ability” of clauses into grammatically legal, information-conveying sentences that Chomsky (1975) was referring as “an intricate structure of specific rules and guiding principles.” Collectively, these rules are referred to as “syntax,” or “grammar.”

Finally, there is *discourse*, which comprises any collection of sentences that can be construed as comprising a unit of communication (e.g., a paragraph, a song, a book, ...).

If we’re interested in understanding how speech is processed differently from nonspeech sounds, it makes sense to start at the most basic level, the phoneme. Studying the perception of words, for example, risks confounding neural signal associated with processing the meaning of the word vs. neural signal associated with detecting and processing a speech sound, per se. Of course, as we’re about to see, limiting oneself to just studying phoneme perception doesn’t mean that one’s task will be easy.

The speech signal

Speech perception presents several difficult engineering problems. One of these has been summarized as the “lack of invariance” problem. That is, to understand spoken language, we need to be able to categorize sounds as corresponding to specific phonemes. However, the same phoneme can often be carried by acoustic patterns that are physically different from one another. When hearing the English words “doll” and “deal” for example, one needs to categorize the initial sound from both words as being the same: /d/. (Mistakenly categorizing one of them as /t/, for example, would result in mistakenly hearing these words as “tall” or as “teal.”) The problem of the lack of invariance, however, is that the acoustic signals when one is saying [da] vs. [di] are quite different (*Figure 18.3*; *Tip 18.3*). Indeed, there are no invariant cues to the identity of any phoneme!

Although lack of invariance can pose problems for the auditory system, it creates opportunities for the scientist. Consider the following. A complement to the property illustrated in *Figure 18.3* is the fact there exist pairs of phonemes whose acoustic properties are very similar, yet they are perceived differently. The most studied examples are phonemes that differ only along one physical dimension, and depending on where along that dimension the sound falls, it is perceived as one or the other. A famous example from English are the phonemes [da] and [ta]. The consonants /d/ and /t/ are from a class of consonants called “stop consonants,” whose utterance requires a momentary blockage of the vocal tract, typically with the tongue. The voice onset timing (VOT) of a stop consonant refers to the latency between the release of the stop and the onset of vibration in the vocal cords. [da] and [ta] differ from one another only in that [da]

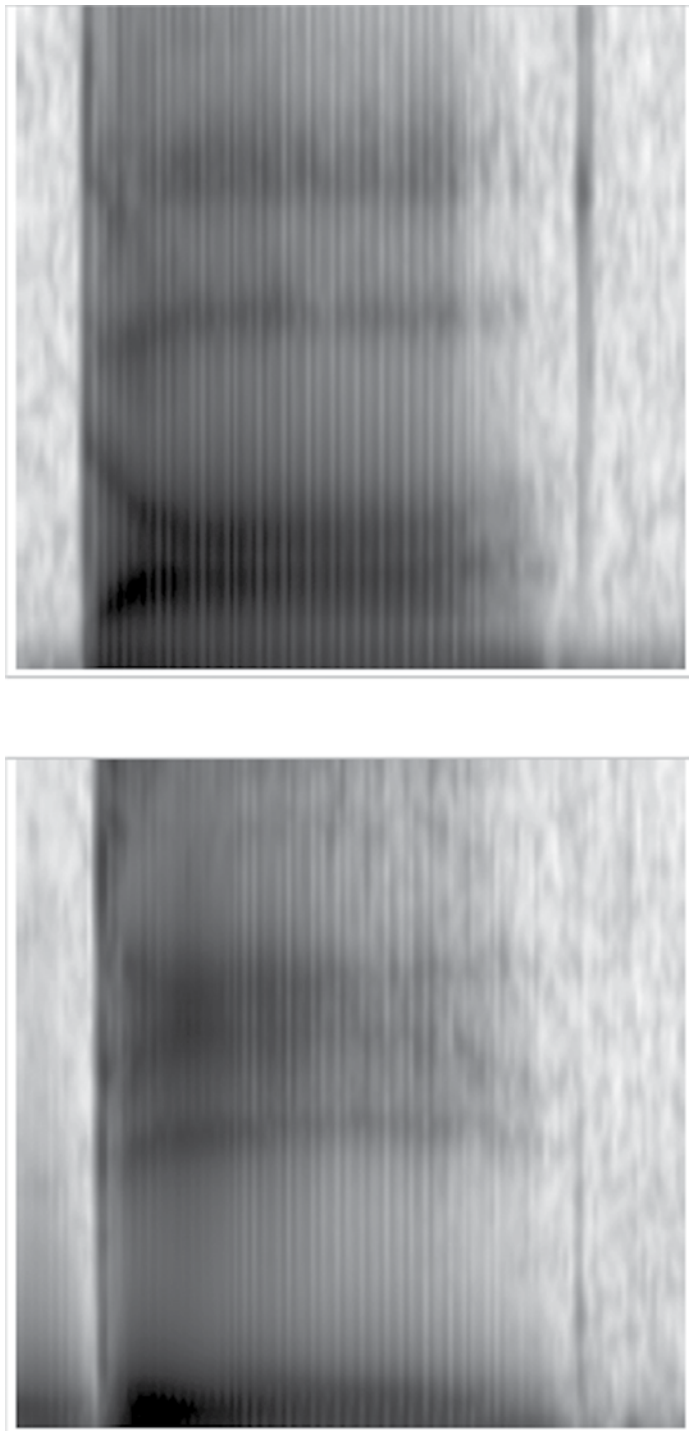


FIGURE 18.3 Spectrograms from an individual uttering the phonemes [da] (top) and [di] (bottom). The vertical axis is frequency (scale of 0–5000 Hz) and the horizontal axis is time (scale of 0–300 ms), with darker gray representing greater power. Note how the early portion of the speech signals differ, even though both are perceived as /d/. The “dark bands” of sustained, elevated oscillatory power that are visibly evident in these spectrograms are called formants. Formants are labeled from low to high frequency, with the first formant also known as the **fundamental frequency** of the tone.

Tip 18.3

One reason for lack of invariance of speech sounds is the phenomenon of coarticulation: the way that we shape and move our speech apparatus to utter the current phoneme is influenced by preparations already underway for the next phoneme to be spoken.

Tip 18.4

Categorical perception of phonemes is one of language's theoretically loaded topics. Whereas some have pointed to it as evidence for a natively endowed speech processor (e.g., Pinker, 1994; Liberman, 1996), others have argued that, because rodents (Kuhl and Miller, 1975) and birds (Kluender, Diehl, and Killeen, 1987) can be taught to make comparable categorical discriminations, it can be explained by learning processes that are common across many vertebrate brains.

has a VOT of 0 ms, and [ta] a VOT of approximately 70 ms. Experimentally, it has been shown that if one begins with [da] and gradually modifies the sound by increasing the VOT from 0 ms out to 70 ms [ta] in, say, 5 ms intervals, subjects do not hear a smooth transition of sounds incrementally changing from [da] to [ta]. Rather, what they hear is a series of undifferentiated “[da]’s until, at one point (typically around a VOT of 35 ms), they suddenly begin hearing a series of undifferentiated “[ta]’s. What this means is that the physical stimulus is changing continuously, but the percept has only two states (*Tip 18.4*). Finding the first station in the auditory system that tracks the percept, as opposed to tracking the physical signal, would be a candidate for the first area that processes speech sounds differently than nonspeech sounds.

Neural correlates of speech perception

In 2005, a team led by the French pediatrician-turned-developmental cognitive neuroscientist, Ghislaine Dehaene-Lambertz, published a study that followed a logic similar to the one just described. Specifically, it employed a mismatch paradigm, which is one of the

most commonly used to study differences in the auditory/language processing system. It is detailed in *Research Spotlight 18.1*.

The ERP findings from Dehaene-Lambertz et al. (2005), indicating that the earliest cortical AEPs discriminate speech- from nonspeech-signals, led them to hypothesize that the two rely on “partially distinct neural processes” (p. 29). Although they don’t speculate about the origin of these processes, they do note that “Further research should determine to what extent this organization is laid down in the course of language acquisition and to what extent it is already present early in infancy” (p. 32). In the ensuing years, Dehaene-Lambertz and colleagues have performed some of this “further research,” with one recent finding, using near-infrared spectroscopy (NIRS; a.k.a. “optical imaging”) on preterm infants (28–32 weeks gestational age) coming down decidedly on the side of “already present” (see *Further Reading* by Mahmoudzadeh et al., 2013; and *Tip 18.5*).

Dual routes for speech processing

One model for speech perception that is generally consistent with the network of regions illustrated in *Figure 18.4.B* is a “dual-route” model advocated by Hickok and Poeppel (2000) whereby “the posterior–superior temporal lobe, bilaterally, is the primary substrate for systems involved in constructing sound-based representations of heard speech” (p. 134), and processing then bifurcates into a ventral and a dorsal stream. The ventral stream is proposed to subserve auditory comprehension via connections that support “interfacing sound-based representations in auditory cortex

Tip 18.5

Near-infrared light passes through the scalp, and different wavelengths are then absorbed by oxygenated vs. deoxygenated hemoglobin, thereby producing a BOLD signal. Although the optics are such that only superficial regions of cortex can be imaged, portability and ease of use relative to fMRI make it a preferred tool for pediatric neuroimaging.

RESEARCH SPOTLIGHT

18.1 ERP and fMRI of speech perception

The stimuli that Dehaene-Lambertz and colleagues (2005) presented to subjects were “synthetic sinewave analogues of speech” that could be interpreted as either whistles or as the phonemes [ba] or [da]. Without prompting, most hear them as whistles. Instead of varying VOT, which differentiates [da] vs. [ta] in English, the continuum along which [ba] vs. [da] differ in French is **place of articulation** – the physical location in the mouth where the tongue applies the blockage that makes the /b/ and /d/ stop consonants. Varying place of articulation influences the acoustic signal by changing the frequency of its **formants** (see *Figure 18.3*). The stimuli were labeled according to their phonemic identity for French speakers and according to the onset frequency of the second formant: ba975, ba1250, da1525, da1800. (Note that, to achieve such precise and replicable acoustic properties in the stimuli, it’s common for speech perception research to use synthesized, rather than natural stimuli.)

Stimuli were presented in groups of four (*Figure 18.4*). In repetition trials, one of the four stimuli was repeated four times. In within-category change (WC) trials the last stimulus differed from the previous three, but still fell on the same side of the phonemic boundary (e.g., *ba975 ba975 ba975 ba1250* or *da1800 da1800 da1800 da1525*). In across-category change (AC) trials the last stimulus fell on the other side of the phonemic boundary (e.g., *da1525 da1525 da1525 ba1250* or *ba1250 ba1250 ba1250 da1525*). Thus, the acoustical distance between the last stimulus and the preceding ones was identical in the two types of change, the only difference being that, when the stimuli were perceived as speech sounds, the change in AC trials would be perceived as a change of phonemic category. This created the critical contrast whereby a change in the auditory percept could be dissociated from changes in the acoustic signal. This comparison was implemented experimentally by performing the experiment in two stages. In “Nonspeech mode,” which always came first, the stimuli were referred to as “whistles,” and subjects were instructed to indicate whether or not the fourth whistle of each trial differed from the others in pitch. Prior to the second stage – “Speech mode” – the speech-like nature of the stimuli was explained to subjects, and they were played the two most distinctively different sounds until they could identify them as [ba] and [da]. Then the procedure was repeated.

Results from the ERP study (*Figure 18.4.A*) indicated that the identical physical change (of 275 Hz) produced significantly greater changes in the ERP for AC vs. WC changes, only in speech mode, beginning with the P1 component. Because the P1 has been localized to parabelt cortex of the STG, this suggests that speech is segregated from nonspeech at a very early stage of cortical processing. When the same experimental procedure was repeated, with different subjects, during fMRI scanning, three patterns emerged (*Figure 18.4.B*): (1) midline and lateral frontal areas, insula, and subcortical regions were sensitive to any change of stimulus; (2) the posterior portion of the superior temporal sulcus was sensitive to mode of perception (i.e., speech > whistle); and (3) supramarginal gyrus was sensitive to a change of phoneme (i.e., AC > WC).

Research Spotlight 18.1 *Continued*

FIGURE 18.4 Dissociating speech perception from nonspeech perception. Source: Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert, and Dehaene, 2005. Reproduced with permission of Elsevier.

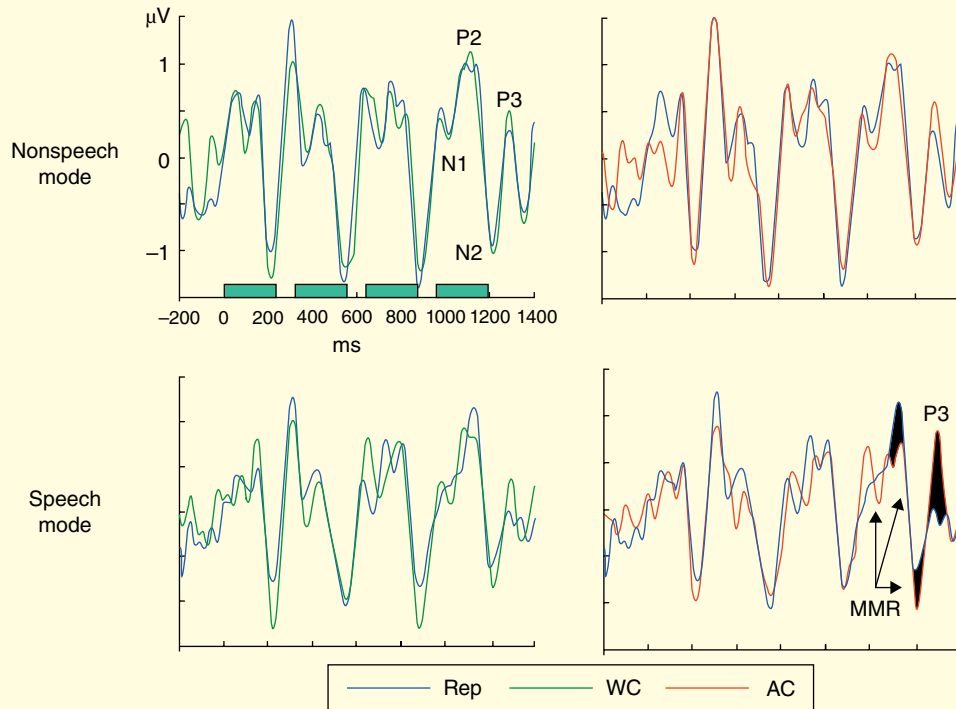


FIGURE 18.4.A ERPs from the Fz electrode for repeat trials (Rep), collapsed across stimuli, plotted with each of the two types of mismatch trials. Along the abscissa of the plot in the upper left, the timing and duration of each stimulus is illustrated with green blocks. The mismatch response (MMR) was prominent, beginning with P1, for stimuli in speech mode that produced an across-category (AC) change.

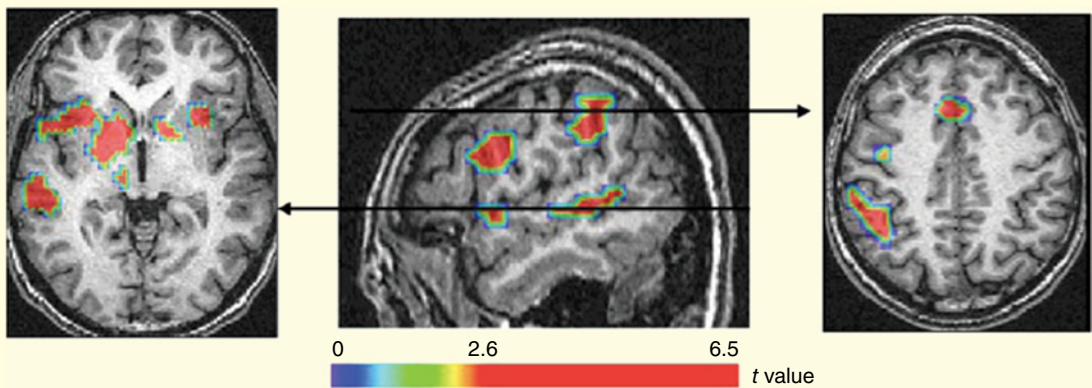


FIGURE 18.4.B Differential sensitivity of these regions to specific contrasts identified with fMRI as showing greater signal for AC vs. Rep in speech mode.

with conceptual–semantic information [represented in] multimodal cortical fields in the vicinity of the left temporal–parietal–occipital junction” (p. 134). Note that these “multimodal cortical fields” correspond to portions of the temporal- and occipital-lobe areas that we studied in considerable detail in *Chapter 12, Semantic Long-Term Memory*. Further, Hickok and Poeppel readily acknowledge that their model draws heavily on ideas first proposed by Wernicke and by Lichtheim. Thus, their ventral stream can be understood as corresponding to the pathway *A–B–M* as portrayed in Figure 18.2.A, and magnified in Figure 18.2.B. (We’ll consider a modern-day depiction of this pathway (Figure 18.6.B) in the upcoming section on grammar.)

The dorsal stream is proposed to act as an “auditory–motor interface” that, in effect, follows the course of the arcuate fasciculus, but with at least some fibers synapsing in inferior parietal cortex. The argument is that posterior temporal regions, as well as inferior parietal regions, contribute to assembling articulatory codes for speech. One consequence of this is the involvement of “dorsal stream” regions, particularly in inferior parietal cortex, in processing phonetic segments, such as stimuli of the kind featured in *Research Spotlight 18.1*.

Note that although *Figure 18.4.B* does not show evidence of elevated right hemisphere activity for the “AC > Rep” contrast illustrated from the Dehaene–Lambertz et al. (2005) study, Hickok and Poeppel (2000) review several neuropsychological and neuroimaging studies that implicate right STG in at least some aspects of speech perception. Although beyond the scope of this book, a comprehensive review of data and theories relating to the lateralization of processing of speech and other acoustical signals is provided in the *Further Reading* from Zatorre and Gandour (2008). Also a *Further Reading* is an alternative dual-route model, from Scott and Johnsrude (2003).

GRAMMAR

Up until this point, we have focused on the perception of speech sounds, and referenced our earlier treatments of single-word comprehension in *Chapter 4* and *Chapter 12*. Now, let’s turn to the decoding of the

language signal. After all, it’s one thing to understand the meaning of the individual words *dog*, *bites*, and *man*, but with this knowledge, alone, one couldn’t know who did what to whom.

Genetics

Here is a partial summary of impairments of four individuals diagnosed with a developmental language disorder:

They took a long time to name pictures of objects with which they were familiar, and tended to use approximate words, for example “glass” or “tea” for cup, and “sky” for star. Comprehension was also delayed, especially the understanding of comparatives. “The knife is longer than the pencil” was poorly understood. “The girl is chased by the horse” became “the girl is chasing the horse,” and “the boy chasing the horse is fat” was interpreted as being the same as “a thin boy chasing a fat horse.” Most of those affected, especially the children, could not retain three items in the correct sequence. Many were unintelligible and reluctant to offer spontaneous conversation. (Hurst, Baraitser, Auger, Graham, and Norell, 1990, p. 354)

The striking thing about these four is that they were all drawn from three generations of the same family (the “KE family,” as Hurst and colleagues refer to them): “The affected members of this family all have the same type of speech and language difficulty, but to varying degrees of severity. The unaffected members have no speech and language difficulties and all attended normal schools” (p. 354). This was a remarkable finding because, as Hurst et al. (1990) noted,

Few speech disorders are inherited in a simple Mendelian way, and when familial clustering does occur, both genetic and environmental factors are involved. The genetic contribution in most cases involves more than one gene (polygenic inheritance), although a dominant gene with reduced penetrance is possible. In practice, it is unusual to find full penetrance of a speech disorder occurring in a regular manner over generations. (p. 352)

Let’s unpack this statement. Inspection of the KE family’s pedigree (*Figure 18.5*) indicates that, beginning with the first generation on record, each child has had a 50% chance of acquiring the disorder. This is a hallmark of autosomal dominant transmission of a mutation in a single gene. That is, in one parent, one

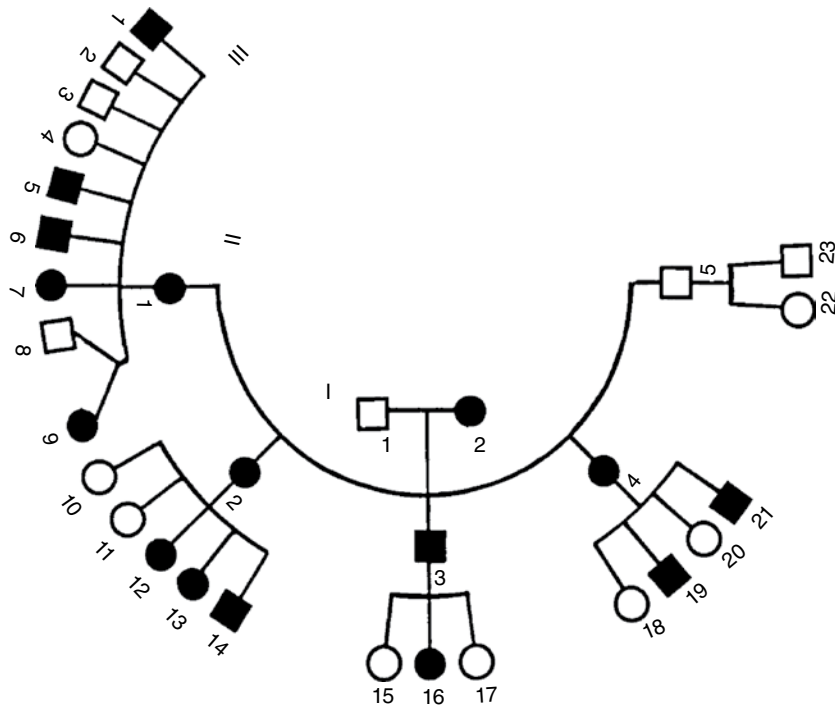


FIGURE 18.5 The pedigree of the KE family. Squares represent males; filled symbols represent affected individuals. Roman numerals identify the three generations of the family, Arabic numerals the pedigree number of an individual within a generation. Source: Hurst, Baraitser, Auger, Graham, and Norell, 1990. Reproduced with permission of Wiley.

of two alleles of this single gene carries a mutation. As a consequence, because the sperm and the egg each contribute one allele from each gene, there is a 50% chance that any child produced by that couple will inherit the mutation. The mutation is dominant, meaning that it determines the transcription and translation of that gene. (For eye color, a brown allele is dominant over blue, such that if a child inherits a blue allele from Mom and a brown allele from Dad, the child's eyes will be brown.) "Penetrance" refers to likelihood that, if one does inherit the mutation, it will be expressed in that individual's phenotype; *Figure 18.5* indicates that the disorder is fully penetrant. Autosomes are all chromosomes that are not the sex chromosomes – i.e., the X and the Y – and therefore an autosomal disorder can be inherited by either male or female offspring. (For a primer on inheritance of single-gene disorders, see *Further Reading* by Chial, 2008.) That's the elementary Mendelian genetics.

What's most striking about this case is that it indicates that a single gene is responsible for the KE family's language disorder!

Autosomal dominant disorders are uncommon. (I just looked up a list of exemplars and the only one I'm confident that I recognize is Huntington's disease.) They are almost unheard of for purely cognitive disorders. Thus, considerable excitement, and controversy, was generated by the linguist Myrna Gopnik, and subsequently others, when she proposed that the function of the mutated gene is to "affect the capacity to acquire language" (Gopnik, 1990b, p. 26). In the interpretation of some (including Pinker, 1994), this was evidence for a "grammar gene." Read about the excitement, and the controversy, in *Thought Box 18.1*, and at *Web Link 18.4*.

The cumulative effect of the findings summarized in *Thought Box 18.1* is that, in this "post FOXP2 era," most would endorse the statement that a complex

THOUGHT BOX

18.1 Is there a “grammar gene”?

The reason the initial summary of KE family deficits was qualified as “partial” is because the initial report from Hurst et al. (1990) also included the fact that, for example, “Articulation was also defective, and [the affected family members had] ... a moderate to severe degree of dyspraxia. They could position the tongue and lips for simple movements, but failed when a sequence of movements was required” (p. 354). These would be considered by many to extend beyond grammar per se. Nonetheless, where grammar ends and where extra-grammatical, but nonetheless “language related,” processes begin is not well defined.

Gopnik’s original (1990a) report on the KE family prompted immediate objections to her interpretation that noted, for example, that the impairments of affected family members were much broader than portrayed, and also included “impairment in repeating phonemes, words, non-words, and sentences... [and in] aspects of language that are unrelated to syntax[, including] semantic errors in naming ... [and poor] receptive vocabulary” (Vargha-Khadem and Passingham, 1990, p. 226). Gopnik’s reply to these objections, in turn, was that

What really is at stake here are not the data, but rather how to account for them. My explanation hypothesizes that the subjects are impaired because they cannot construct language rules at the necessary level of abstraction. [Fletcher (1990) and Vargha-Khadem and Passingham (1990)] prefer an explanation which hypothesizes that the subjects are impaired because they cannot hear or cannot pronounce certain sounds, and this indirectly affects the grammar carried by these sounds. (Gopnik, 1990b, p. 26)

When confronted with such a case of “he said-she said,” what is one to do? The answer, of course, is that one is to perform studies that are carefully designed to assess the key questions of the debate. This additional research proceeded along two paths, one behavioral, one genetic. The behavioral, pursued by Vargha-Khadem, Watkins, Alcock, Fletcher, and Passingham (1995), elaborated on the arguments that they had raised in their initial response, and also established that affected members had significantly lower verbal and nonverbal IQ, with the score of many being below the value “commonly considered an exclusionary criterion for classification of a subject as having a specific language impairment” (p. 932).

Perhaps the most compelling evidence against the “grammar gene” position, however, arrived with the identification of the gene itself. In 2001, geneticists Cecilia Lai and Simon Fisher, working in the laboratory of Anthony Monaco at Oxford University, reported the identification of the mutated gene as FOXP2. This was an important breakthrough, because the functions of the gene could then be studied, independent of the KE family or others carrying the mutation. From research that both preceded and followed the report of Lai, Fisher, Hurst, Vargha-Khadem, and Monaco (2001), we know that FOXP2 acts as a regulator (primarily a repressor) of the transcription of other genes. Graham and Fisher (2013) have noted that “FoxP2 is likely to be present in all vertebrates, and is highly conserved in neural expression pattern and amino-acid coding sequence ... Thus, ancestral versions contributed to brain development long before language appeared” (p. 44). In mice, the mutation is associated with impaired motor learning and electrophysiological abnormalities in the striatum. (See *Web link 18.4*.)

cognitive faculty like grammar is unlikely to be isolatable to a single gene. At a different level, another lesson that one can take away from the saga of FOXP2 is that, for some developmental disabilities, it is not uncommon for there to be disagreement about what constitutes a fundamentally linguistic impairment vs. a more general cognitive impairment. (This is also seen, for example, with so-called specific language impairment (SLI) and with dyslexia, which will be the focus, further along, of *Real-World Applications 18.1*.) If we want to address the cognitive neuroscience of grammar head-on, then, let's move to a different literature.

Rules in the brain?

Grammar refers to the systematic ways by which one can combine and modify words to encode (if speaking) or decode (if listening) a precise message. One property of grammar is that it is combinatorial. For example, the phrase structure of grammar allows one to go from "The boy is fat" to "The boy chasing the horse is fat," to "The boy chasing the horse on the reduced-calorie diet is fat," with nary any uncertainty about which organism's body type is being called out. Another property of grammar is that it allows one to convey temporal context by changing verb tense: "The boy was fat" vs. "The boy is fat" vs. "The boy will be fat." These are descriptions of grammar that no one would contest as untrue. Where there is contention, however, is whether these operations – combinatorial recursion and verb inflection – depend on rules.

Pinker and Ullman (2002a) have summarized it as follows:

The [English] past tense is of theoretical interest because it embraces two strikingly different phenomena. Regular inflection, as in walk-walked and play-played, applies predictably to thousands of verbs and is productively generalized to neologisms such as spam-spammed and mosh-moshed, even by preschool children. Irregular inflection, as in come-came and feel-felt, applies in unpredictable ways to some 180 verbs, and is seldom generalized; rather, the regular suffix is often overgeneralized by children to these irregular forms, as in holded and brokeed. A simple explanation is that irregular forms must be stored in memory, whereas regular forms can be generated by a rule that suffixes -ed to the stem. (p. 456)

Indeed, they go on to argue that "rules are indispensable for explaining the past tense, and by extension, language and cognitive processes" (p. 456). Although Pinker, a cognitive psychologist/psycholinguist (we'll define this in a later section), originally made this argument on theoretical and behavioral grounds, his former student/now collaborator Michael Ullman added a neural implementation of this "Words-and-Rules" theory: whereas the past-tense forms of irregular verbs are stored in lexical memory (i.e., posterior temporoparietal cortex), past tense forms of regular verbs are computed online by a concatenation rule ("add -ed") implemented by the cortico-basal ganglia loop that includes Broca's area. Evidence that Ullman and colleagues have marshaled in support of this model has included two demonstrations of neuropsychological double dissociations. In one, a patient with a nonfluent (for our purposes, Broca's) aphasia was shown to have a selective impairment at generating the past tense with regular verbs, whereas a patient with a fluent (for our purposes, Wernicke's) aphasia had a selective impairment with irregular verbs. In a second, patients with Parkinson's disease, which is characterized by progressive deafferentation of the neostriatum from its supply of midbrain dopamine, and thus by progressive striatal dysfunction, have selective difficulty inflecting regular verbs, whereas patients with Alzheimer's disease, associated with episodic and semantic memory loss, have selective difficulty inflecting irregular verbs.

An alternative view has been advanced by McClelland and colleagues. For example, McClelland and Patterson (2002a), in response to Pinker and Ullman (2002a), wrote that the parallel distributed processing (PDP) framework holds that "Acquisition of language and other abilities occurs via gradual adjustment of the connections among simple processing units. Characterizations of performance as 'rule-governed' are viewed as approximate descriptions of patterns of language use; no actual rules operate in the processing of language" (p. 465). Although much of their argument derives from computational modeling, McClelland and Patterson (2002a) also appeal to neural data as consistent with their perspective. For example, they note that among the findings of Vargha-Khadem et al. (1995) was the fact that affected members of the KE family (from *Thought Box 18.1*) showed comparable

levels of impaired past-tense processing of regular and irregular verbs. Further, with Bird, Lambon Ralph, Seidenberg, McClelland, and Patterson (2003), they have argued that a key difference between regular and irregular verbs in English is that “relative to most irregular past-tense forms, regular past-tense English verbs place greater demands on the phonological system, because the transformation from regular stem to past always involves the addition of extra phonemes whereas this is rarely the case for irregular verbs” (p. 503). They suggest that some studies, including the double dissociations invoked by Pinker and Ullman (2002a), may thus be explainable in terms of patients with anterior lesions having more difficulty with items with greater phonological complexity. Further, Bird et al. (2003) demonstrated in 10 nonfluent (i.e., Broca’s) aphasics that “the apparent disadvantage for the production of regular past tense forms disappeared when phonological complexity was controlled” (p. 502).

“Immediate” replies to the two positions summarized here can be found at Pinker and Ullman (2002b) and McClelland and Patterson (2002b), and in many ensuing publications. The debate about whether the brain truly represents rules, or, alternatively, rule-like patterns are emergent from complex network dynamics, seems unlikely to end in a consensus agreement during the career of this author. In the meantime, with this context, let’s consider some perspectives on the functions of Broca’s area.

Broca’s area

The classical view from behavioral neurology, as embodied by the Wernicke–Lichtheim model, is that Broca’s area is critical for syntactic processing – that is, for the grammatical decoding of heard sentences, as well as the grammatical encoding of spoken sentences. For example, patients with Broca’s aphasia have difficulty understanding sentences like “It was the squirrel that the raccoon chased,” from which meaning can’t be derived just from knowledge of the meaning of the individual words. Despite this fact, however, others have shown that even though these patients don’t understand these sentences, they can make correct grammaticality judgments about them, suggesting that at least some grammatical knowledge may reside elsewhere in the brain. And so, what is the function of Broca’s area?

The anatomy and connectivity of Broca’s area

Figure 18.6 illustrates the three subregions of what might be referred to as “Broca’s area +,” the two lateral regions that traditionally make up Broca’s area – the inferior frontal gyrus (IFG) pars triangularis (PT_r) and the IFG pars opercularis (PO) – plus the functionally related frontal operculum (FO). Figure 18.6.B schematizes the evidence for dorsal and ventral pathways connecting Broca’s area with posterior temporal cortex – the dual routes referred to earlier in the context of Hickok and Poeppel’s (2000) model of speech perception. The heterogeneity of inputs to subregions of Broca’s area is important for some models that emphasize a heterogeneity of functions within this area.

Functional accounts of Broca’s area

Readers who have made it this far through the book won’t be surprised to learn that the classical view of Broca’s area as the neural seat of grammar has, over the past quarter century, been called into question by many investigators. There’s not room here to detail every alternative account, but some prominent ones are verbal short-term memory (STM), cognitive control, and “unification,” a hypothesized process of integrating information from phonological, lexical, syntactic, and semantic levels in the service of language comprehension and production (Hagoort, 2005; this concept will also be relevant for the

FIGURE 18.6 The anatomy and connectivity of Broca’s area. Source: Rogalsky and Hickok, 2011. Reproduced with permission of The MIT Press.

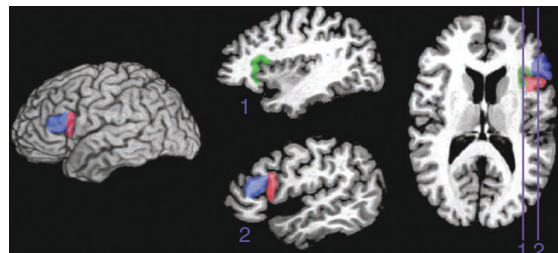


FIGURE 18.6.A The two primary anatomical subdivisions of Broca’s area, plus the frontal operculum. At left is a lateral view of a 3-D-reconstructed brain. The three slices are illustrated in radiological conventions. Blue = pars triangularis; red = pars opercularis; and green = frontal operculum.

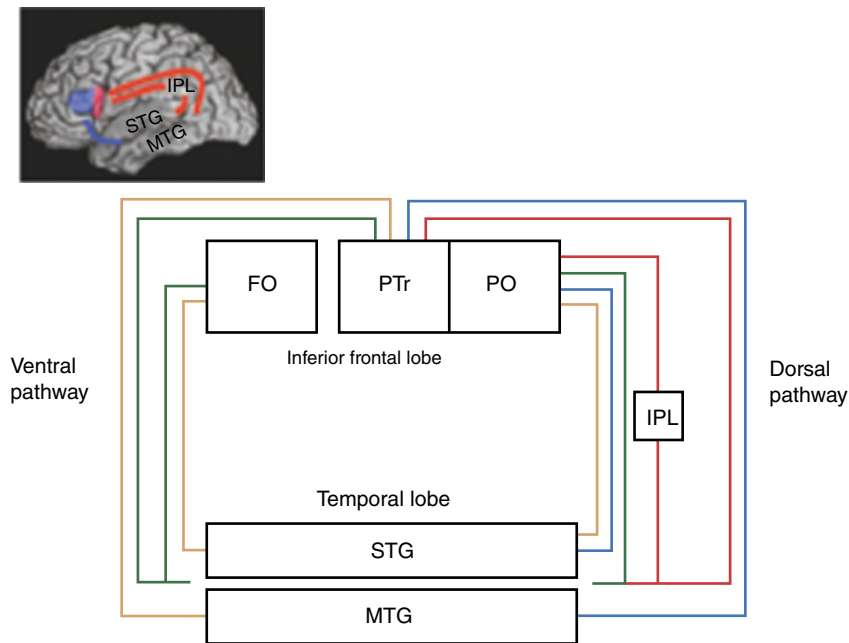


FIGURE 18.6.B A summary of findings from four diffusion tensor imaging tractography studies. The illustration in the upper left summarizes the dorsal and ventral pathways. The box-and-arrow diagram schematizes the findings from four studies (each color coded). STG=superior temporal gyrus; MTG=middle temporal gyrus; IPL=inferior parietal lobule; PO=pars opercularis; PTr=pars triangularis; FO=frontal operculum.

upcoming section on *The electrophysiology of grammar*). Although these alternative accounts differ in terms of the extent to which each may or may not generalize to domains of cognition outside of language, what they have in common is that they propose non-syntactic functions for this region. Two responses to these non-syntactic models have argued that there may be a finer-grained functional organization of Broca's area, such that some subregions do, indeed, support syntactic functions, and others may support non-syntactic ones. This style of reasoning is similar in spirit to one that we saw in relation to the organization of theory-of-mind (ToM) mentalization in the region of the right temporoparietal junction (TPJ), in *Chapter 16*. Indeed, one of the proponents of one "subregional specialization" model of Broca's area is none other than Nancy Kanwisher, whose intellectual fingerprints are also on the ToM debate, in addition, of course, to that of face processing in ventral temporal cortex. Fedorenko, Duncan, and Kanwisher (2012) summarize what's at stake in this debate as follows:

In 1861, Paul Broca stood up before the Anthropological Society of Paris and announced that the left frontal lobe was the seat of speech. Ever since, Broca's eponymous brain region has served as a primary battleground for one of the central debates in the science of the mind and brain: Is human cognition produced by highly specialized brain regions, each conducting a specific mental process, or instead by more general-purpose brain mechanisms, each broadly engaged in a wide range of cognitive tasks? (p. 2059) (*Tip 18.6.*)

Tip 18.6

Lest one think that Fedorenko and colleagues (2012) are unique in invoking martial rhetoric in reference to Broca's area, theirs was preceded by papers entitled, for example, "The Battle for Broca's Region" (Grodzinsky and Santi, 2008), and "Broca's Region: Battles are not Won by Ignoring Half of the Facts" (Willems and Hagoort, 2009).

Two “subregional specialization” models of Broca’s area, defined with univariate analyses of fMRI data, are illustrated in *Figure 18.7*.

Although there are important historical and theoretical reasons for focusing attention on Broca’s area, let’s not lose sight of the importance of network-level activity in supporting language functions. One fMRI study by Abrams et al. (2013), for example, has contrasted activity associated with listening to

snippets from famous speeches from twentieth-century American history vs. listening to unintelligible versions of these same stimuli. (Both sets of clips were digitally filtered to equate them in terms of acoustic properties, but the unintelligible clips were rendered thus by being “spectrally rotated”; see Abrams et al. (2013) for details about how this was done.) The authors reasoned that “Intelligible speech is hypothesized to recruit brain structures supporting

FIGURE 18.7 Two models advocating a functional compartmentalization of Broca’s area.

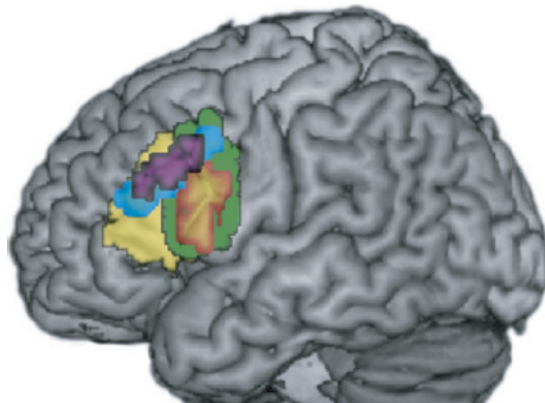


FIGURE 18.7.A fMRI evidence for subregions of Broca’s area sensitive to syntactic complexity (orange) vs. working memory load (teal) of sentences. Syntactic complexity was manipulated by varying hierarchical structure, such as presence of center-embedded clauses, in sentences. Working memory was manipulated by the number of words separating a subject noun from its respective verb, and the region sensitive to this contrast is illustrated in teal. Regions in purple showed elevated connectivity with PO during more syntactically demanding sentence processing, and are interpreted as evidence for increased working-memory support for difficult sentence processing. Yellow is anatomically identified BA 45/PT_r; green is anatomically identified BA 44/PO. Source: Makuuchi, Bahlmann, Anwender, and Friederici, 2009. Reproduced with permission of PNAS Group.

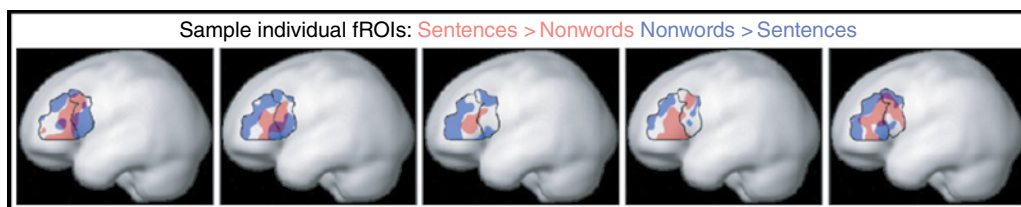


FIGURE 18.7.B Data from Broca’s area of five individual subjects, with black outlines corresponding to BA 45/PT_r and Ba 44/PO. Regions in red responded with greater fMRI signal intensity during sentence-reading followed by a single-word recognition probe, and regions in blue during nonword viewing followed by a single-nonword recognition probe. Subsequently, the sentence-sensitive regions were shown to respond only minimally to tests of mental arithmetic, spatial and verbal STM, and three tests of cognitive control, whereas the nonword-sensitive regions responded with significantly greater signal intensity to difficult vs. easy versions of many of these tasks. Source: Fedorenko, Duncan, and Kanwisher, 2012. Reproduced with permission of Elsevier.

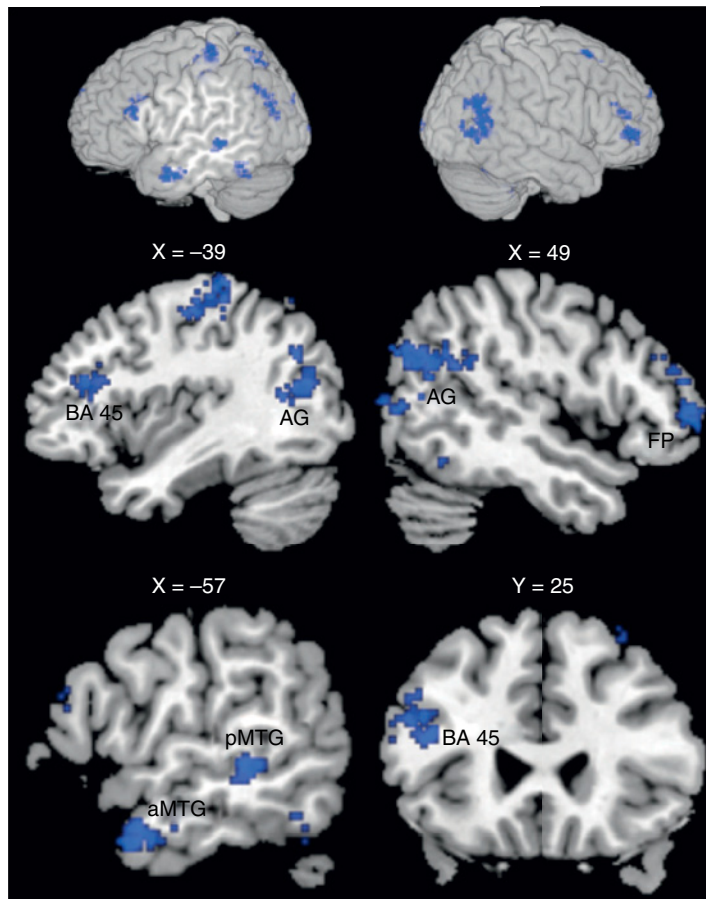


FIGURE 18.8 Regions identified by whole-brain, searchlight-based MVPA as discriminating intelligible from unintelligible sentences. AG, angular gyrus; FP, frontal pole; aMTG, anterior middle temporal gyrus; pMTG, posterior middle temporal gyrus. The left-hemisphere regions identified here were also found to have elevated functional connectivity for intelligible vs. unintelligible sentences. Source: Abrams, Ryali, Chen, Balaban, Levitin and Menon, 2013. Reproduced with permission of Oxford University Press.

phonology, word-form recognition, semantics, and syntax ... while carefully controlled unintelligible speech retains many speech-like qualities ... including frequency and amplitude modulations and formant-like features, but is not thought to provide direct access to linguistic or semantic representations in the brain” (p. 1703). The findings from a whole-brain multivariate pattern analysis (MVPA) of their data, illustrated in *Figure 18.8*, confirm that Broca’s area (more specifically, BA 45/PT_r) is one among many bilaterally distributed areas whose activity discriminates intelligible from unintelligible sentences.

The electrophysiology of grammar

Explicit in the arrows drawn onto Lichtheim’s house model (*Figure 18.2.A*), as well as in many more recent accounts, is the assumption that sentence processing is carried out in a feed-forward (or “bottom-up”) manner: the first stage is the identification of phonemes; followed by word recognition; followed by syntactic decoding of the spoken utterance. ERP studies of sentence processing, however, have provided evidence for high levels of interactivity within the system. For example, *Figure 18.9* illustrates the results from a study

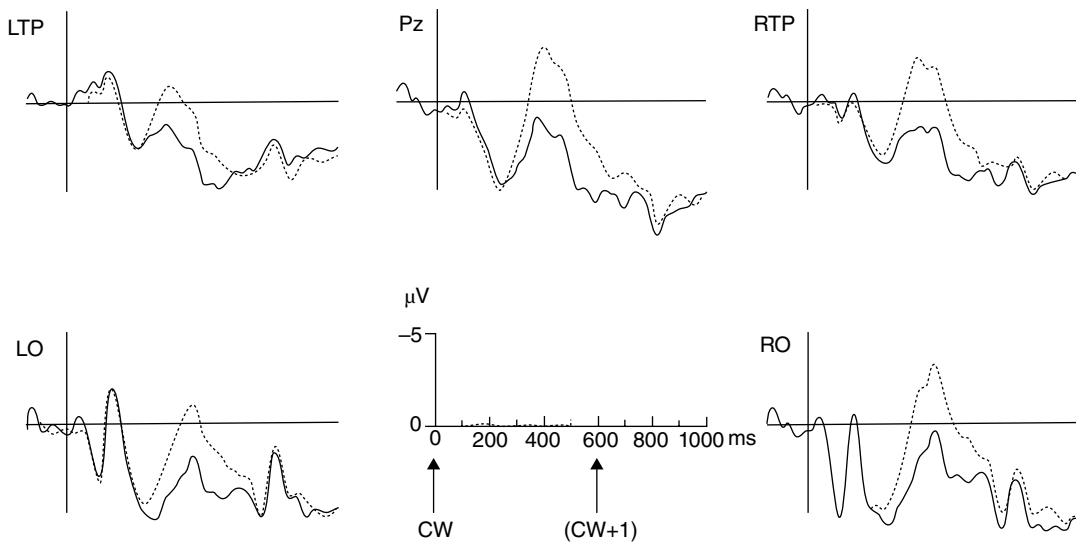


FIGURE 18.9 ERPs time-locked to a critical word (CW) that either was (dotted waveform) or was not (solid waveform) anomalous in the context of the story in which the sentence was embedded. The N400 evoked by the CW is markedly larger in the anomalous condition. Although not shown here, this was true at all scalp locations. (In some sentences the CW was not the sentence-final word, in which case “CW + 1” followed 600 ms later.) Source: Van Berkum, Hagoort, and Brown, 1999. Reproduced with permission of The MIT Press.

by van Berkum, Hagoort, and Brown (1999) in which subjects listened to a series of brief vignettes. The final sentence of each one contained a critical word that was always semantically plausible and always syntactically correct within the sentence in which it appeared, but that could be either congruent or incongruent with the discourse-level context. For example, one vignette (as translated from the Dutch) began with: “As agreed upon, Jane was to wake her sister and her brother at five o’clock in the morning. But the sister had already washed herself, and the brother had even got dressed.” For the critical sentence, subjects then heard “Jane told the brother that he was exceptionally _____,” with the final word being either “quick” (i.e., congruent) or “slow” (i.e., incongruent).

The dependent measure of interest in this study was the N400, an ERP component previously shown to be sensitive to the processing of meaning in several contexts, including spoken, written, and signed language, mathematical symbols, and videos with social content. As the name implies, it typically peaks later than ERP components associated with sensory and perceptual processing (e.g., *Figure 18.4.A*). In this study, evidence that contextually incongruent words

are differentially processed seems to become evident as early as 200–300 ms after word onset. Kutas and Federmeier (2011) interpret these findings, and those from similar studies, as follows: “Because, in the absence of prediction, these [critical] words constitute equally good fits to the accrued contextual information, N400 reductions when the words matched as opposed to mismatched the predicted target showed clearly that information about likely upcoming words has shaped the system in advance” (p. 634). That is, one perspective on the grammatical decoding of language is to understand it as another example of the principle of predictive coding, which we first introduced at the end of *Chapter 3*. This perspective will also carry through to the final section of this chapter, on speech production. (See *Web Clip 18.1* for more on the N400 as a tool for studying sentence processing.)

SPEECH PRODUCTION

The physical production of speech is an exquisitely complicated motor control problem. We have seen this, for example, in the fact that an offset of a few tens

of milliseconds in the coordination of the release of the tongue from the palate relative to the contraction of the diaphragm to force air over the vocal cords can make the difference in whether one produces the word “doll” or “tall.” Further, we saw with the KE family that dysfunction in the control of the articulatory apparatus can produce profound impairments in the ability to communicate. Our focus here, however, will be on the neural control of speech production. Recall that with Broca’s aphasia, for example, the patients are no better able to produce language with a pen, or a keyboard, than with their voice. Thus, there’s a level of language production that is abstracted from the effectors that will actually generate the output. (Having said this, however, it will be most expedient in this brief chapter to stick with spoken language.)

A psycholinguistic model of production

Within the study of language, a distinction is often made between the fields of *linguistics* and of *psycholinguistics*. Although no single definition will please everyone, it’s a fair generalization to say that linguistics concerns itself primarily with computational and theoretical questions about the structures and rules of language, whereas psycholinguistics addresses how humans process and use language in relation to behavior. As a general rule, the discipline of psycholinguistics tends to be more associated with designing and carrying out experiments. Almost all the topics that we’ve covered in this chapter fall under the rubric of psycholinguistics.

According to one influential model, speech production occurs via a series of stages, beginning with the formulation of a message, followed by lexical retrieval (mapping an abstract, conceptual message onto words), phonological encoding (specifying the speech sounds that comprise the words), articulatory planning (formulation of a sequential motor plan for speech output), and articulation (Indefrey and Levelt, 2004). It is likely that many of these processes recruit systems that we have already considered. For example, repetitive transcranial magnetic stimulation (rTMS) of left posterior STG, timed to interfere with the phonological encoding stage of production, does, indeed, produce errors in spoken output (Acheson, Hamidi, Binder, and Postle, 2011). The area targeted for rTMS

in this study overlapped with the region of STG highlighted in *Figure 18.4.B*, and the processes disrupted by rTMS likely overlapped with those discussed earlier in the context of articulatory coding functions of the speech perception system (i.e., the “dorsal stream”).

Forward models for the control of production

Decades of psycholinguistic research on speech production indicate that it, too, is a highly dynamic and interactive process. In particular, our production of speech is very sensitive to our online assessment of it. Here’s a summary from Houde, Nagarajan, Sekihara, and Merzenich (2002):

Adding noise to a speaker’s auditory feedback results in an elevation of voice volume ... Delaying a speaker’s auditory feedback by more than about 50 msec produces noticeable disruptions in speech ... Shifting the spectrum of auditory feedback causes shifts in the spectrum of produced speech ... Perturbing perceived pitch causes the subject to alter his pitch to compensate for those perturbations ... Altering perceived formants induces compensating changes in the production of vowels. (p. 1125)

These findings have given rise to the idea that speech production may involve the construction of a neural **forward model** of the planned utterance, against which the actual speech output is compared. Recall that we have already seen the same principle at work in *Chapter 7*, in the context of the “internal models” of movements represented in the cerebellum, and their use in fine-tuning skeletomotor control. (For our purposes, the terms “forward model” and “internal model” are interchangeable.)

To test this idea, Houde, Nagarajan et al. (2002) recorded the magnetoencephalogram (MEG) while subjects uttered a single vowel and heard its real-time playback, and when they passively listened to the acoustically identical signal (i.e., to a recording of themselves from the speaking session). The data, illustrated in *Figure 18.10.A*, showed a marked attenuation of the auditory evoked response in the speaking condition, which the authors attributed to “a comparison between the incoming signal and an internally generated prediction of what that signal would be” (p. 1131). When the match was high, they reasoned, the “error”

would be minimal, and so, too, would be the neural response. To rule out any possible confounds due to the fact that one condition involved actual production and one did not, they performed a control study that was identical to what was just described, with the exception that the auditory feedback of their voice was convolved with white noise. Thus, the auditory signal was perfectly matched to the timing of hearing one's own voice, but was a signal that the auditory system could not predict that it would hear. The result, comparable responses in both conditions (*Figure 18.10.B*) ruled

out many confounding alternative explanations to the “forward model” interpretation. (Conceptually similar results have also been demonstrated with fMRI, implicating bilateral STG in processing mismatches between expected and actual auditory signals (Tourville, Reilly, and Guenther, 2008).)

Before moving on, let's take a moment to consider the potentially very profound implications of these findings and their interpretation. One construal of auditory evoked responses is that they correspond to (they “signal,” if you will) mismatches between the

FIGURE 18.10 MEG responses to hearing one's own speech. Source: Houde, Nagarajan, Sekihara, and Merzenich, 2002. Reproduced with permission of The MIT Press.

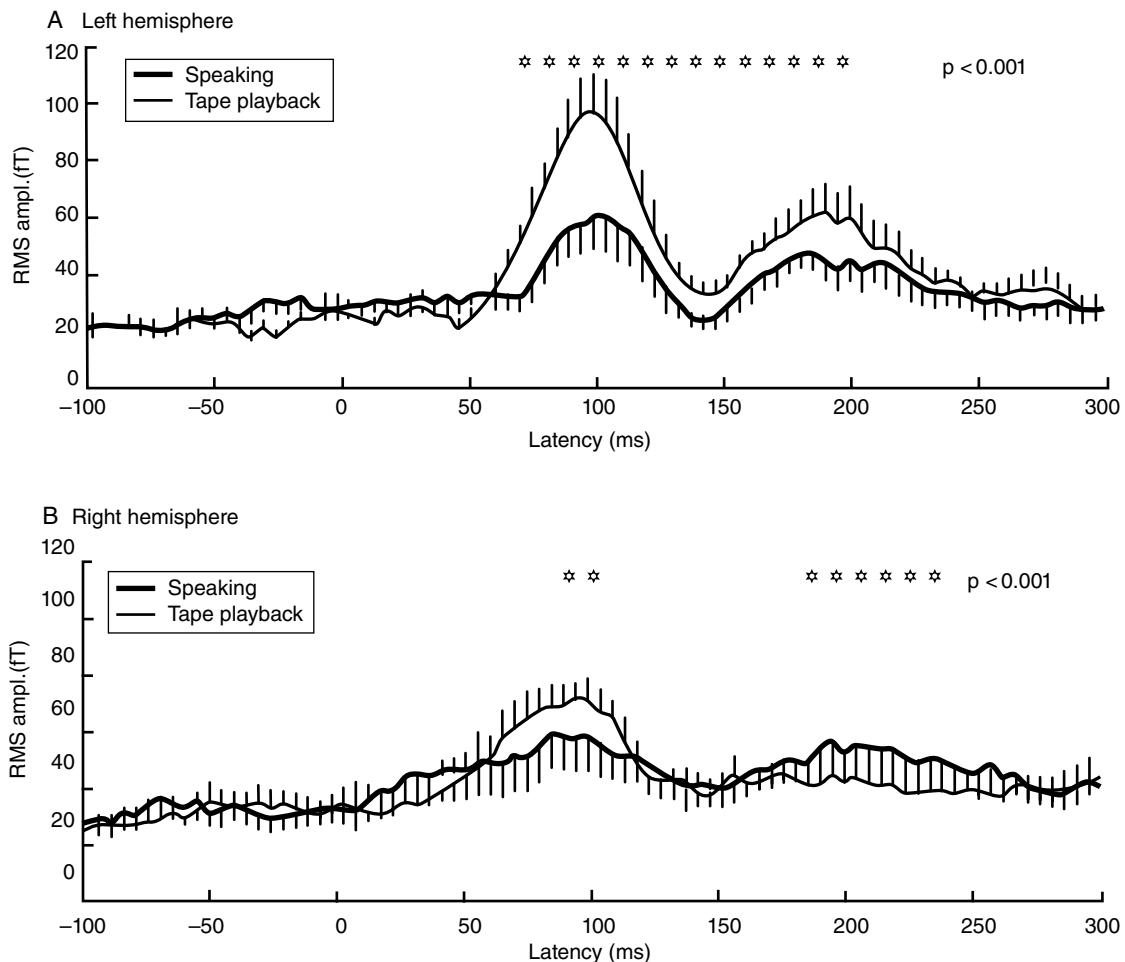


FIGURE 18.10.A MEG evoked response, time-locked to sound onset, from subjects hearing their own utterance of a vowel while speaking it (“Speaking”) and while passively listening to it “Tape playback”. The response is significantly attenuated in the speaking condition.

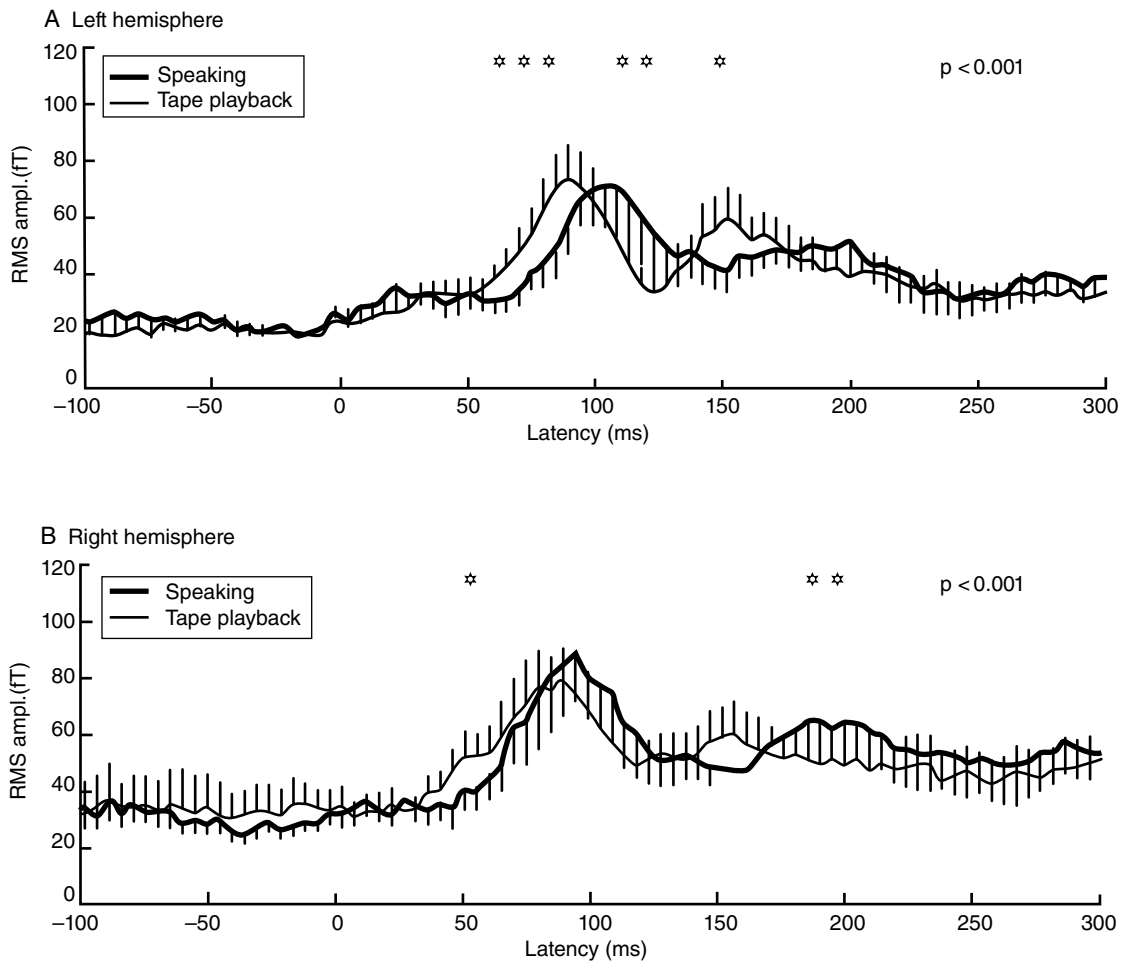


FIGURE 18.10.B MEG evoked response, time-locked to sound onset, from subjects hearing a burst of white noise of the same duration of their utterance of a vowel, while speaking it (dark line) and while passively listening to it. With noise, the evoked responses are the same magnitude during Speaking vs. Tape playback.

brain's expectation about the environment and the actual state of the environment. In relation to this section of the chapter, the studies reviewed here suggest that production is a highly interactive process that recruits many levels of neural representation of sound, meaning, and syntax.

Necessary neural substrates?

Finally, although we've noted that various posterior regions seem to make important contributions to

speech production, let's return to Broca's area. To riff on the martial metaphor, might an analogy between the "Battle for Broca's area" and the D-Day (June 6, 1944) assault by US Army Rangers on the Pointe du Hoc be apt? (After fierce fighting and heavy losses, the US soldiers took the position, only to find that the objective of their attack, large artillery guns, had been disassembled and moved days earlier.) *Methodology Box 18.1* offers another of our periodic reminders about the importance of studies that support causal inference.

METHODOLOGY BOX

18.1 Voxel-based lesion-symptom mapping (VLSM) of speech production and comprehension

Traditional approaches to neuropsychology, although valuable in many instances, can be susceptible to bias. One is to group patients by locus of lesion. A potential source of bias with this method is that the existence of multiple functional subregions may be obscured if many of the patients have lesions that span these subregions. (An analogous caveat about TMS was raised in *Chapter 16*'s treatment of the study by Young, Camprodon, Hauser, Pascual-Leone, and Saxe, 2010.) A second traditional method is to select patients based on their symptoms; after identifying individuals with similar behavioral impairments, one looks for a locus of lesion overlap that might explain the syndrome. The potential bias with this approach relates to specificity: when employing a symptom-based approach one necessarily only selects patients whose behavior manifests the symptom(s) that one is interested in. Thus, even if area X is damaged in each patient in one's sample, one cannot know whether, in recruiting patients, one may have overlooked a comparable number who also have lesions of area X, but who don't manifest the symptom(s) of interest.

VLSM avoids these two sources of bias by allowing for the study of deficit-lesion correlations without requiring the categorization of patients, either in terms of lesion location or in terms of behavior. Rather, both can be treated as continuous variables, and the procedure evaluates the relation of tissue integrity to behavior on a voxel-by-voxel basis. Thus, it is effectively the same approach that is used in massively parallel univariate analysis of functional neuroimaging data, the difference being that there is no variation in the neural data over time.

A study by Bates and colleagues (2003) applied VLSM to 101 patients who had suffered left-hemisphere strokes and showed some degree of consequent language impairment. Behavioral measures of verbal fluency and auditory comprehension were acquired from subtests of a standard clinical assessment tool, the Western Aphasia Battery. Each patient's fluency score summarized performance on articulatory, word-finding, and sentence-production tasks, and the auditory comprehension score summarized performance on yes/no questions, single-word recognition, and enactment of one-, two- and three-part commands. Lesion maps were transformed into a common space, and each voxel then binarized for each subject as "intact" or "lesioned." (That is, tissue integrity was not treated as a continuous variable in this study, although one could do as with, for example, voxel-based morphometry (VBM) data.)

The results, illustrated in *Figure 18.11*, suggest a more prominent role for insular cortex than Broca's area for fluency, and a more prominent role for middle temporal gyrus than superior temporal gyrus for comprehension. For arguments supporting these two findings, see Dronkers (2000) and Binder (2003) in the list of *Further Reading*.

A caveat not unique to VLSM, but salient here, is that stroke damage is often correlated across regions, because one arterial branch typically feeds multiple regions. Thus, for example, it could be that the identification of the superior longitudinal fasciculus, which primarily connects posterior parietal cortex with PFC, is incidental to concurrent damage to the arcuate fasciculus, which actually does contribute importantly to production.

(Note that we also saw this approach in *Figure 6.3*, where it was used to determine regions whose damage was associated with acute vs. chronic unilateral neglect.)

Methodology Box 18.1 *Continued*

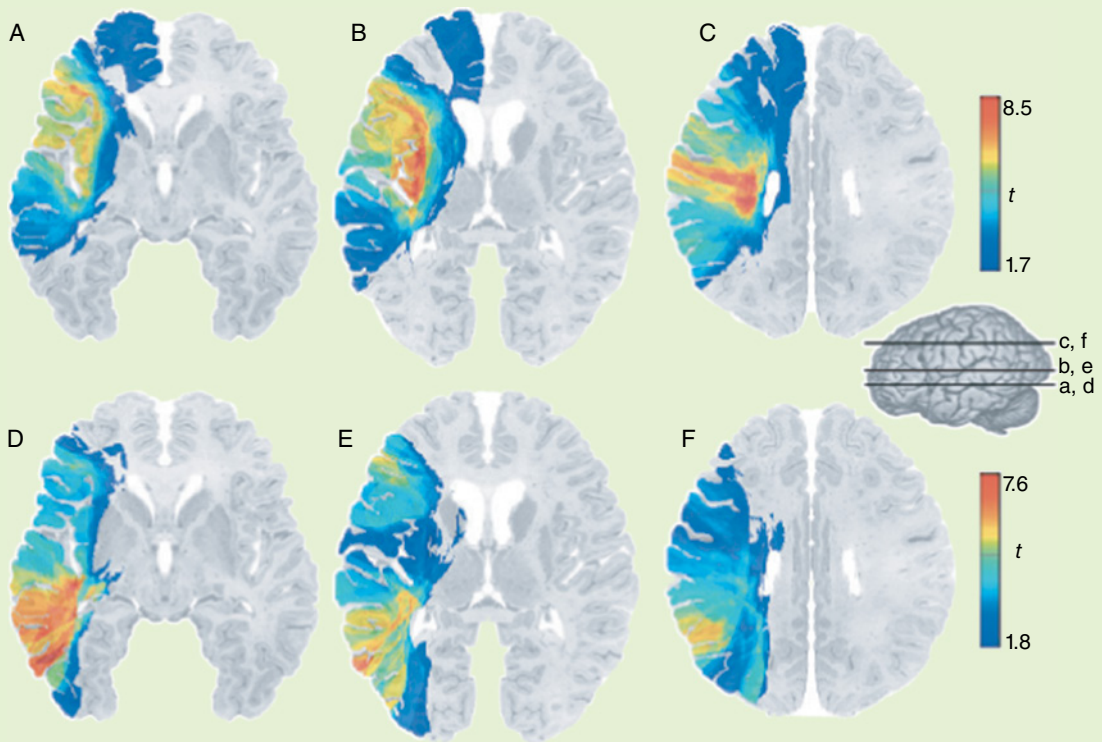


FIGURE 18.11 Voxel-based lesion-symptom maps (VLSM) from 101 left-hemisphere stroke patients. The top row illustrates the voxel-by-voxel association between lesion status and verbal fluency performance, and the lower between lesion status and auditory comprehension. The color at each voxel corresponds to its t -value for that contrast, just as it does for univariate functional imaging maps. Thus, warmer colors indicate regions where lesions are more strongly associated with behavioral impairment. The faintly visible gray line on each color bar indicates the Bonferroni-corrected significance cutoffs (in the teal range for fluency, and near the teal–yellow transition for auditory comprehension). These analyses suggest that lesions of Broca’s area (**A**) insula (**B**) and white matter encompassing both the arcuate and superior longitudinal fasciculi (**C**) had the greatest impact on fluency, whereas lesions of the middle (**D**) and superior (**E**) temporal gyri had the greatest impact on auditory comprehension. Source: Bates, Wilson, Saygin, Dick, Sereno, Knight, and Dronkers, 2003. Reproduced with permission of Nature Publishing Group.

REAL-WORLD APPLICATIONS

18.1 The cognitive neuroscience of dyslexia

Dyslexia, a persistent difficulty in learning to read that is not explained by sensory deficits, cognitive deficits, lack of motivation, or lack of adequate reading instruction, is estimated to affect between 5% and 17% of children. As noted in the quote from Chomsky that opened this chapter, most children, including those with dyslexia, acquire facility with spoken language with seemingly magical ease. Reading, however, is a different story. Reading must be learned via effortful, persistent instruction.

Dyslexia is strongly heritable, occurring in up to 68% of identical twins and 50% of individuals who have an affected parent or sibling, and has been associated with genes involved in neuronal migration during development. Although dyslexia manifests as a difficulty with an overtly visual behavior, it is widely viewed as being rooted in impaired phonological awareness. For example, children with dyslexia typically have difficulty indicating which of a set of words begins with the same sound as a target word, or being able to break a word into its constituent phonemes. Evidently, the impairment of an explicit understanding of phonology translates into a difficulty mapping individual written letters, or groups of letters, into the sounds that they are meant to represent.

In the brain, diffusion tensor imaging studies have identified differences in white matter integrity both in left posterior temporoparietal cortex, and in fiber tracts connecting this region with anterior and subcortical areas. There is less consensus about the neurophysiological bases of dyslexia. In the 1990s, research provided behavioral and postmortem histological evidence implicating selective impairment in the magnocellular channel of the visual system. The idea was that impaired processing by this fast channel resulted in subtle disruptions of timing between visual and auditory/phonological signals that interfered with reading. Some more recent studies, however, have failed to support this idea. Other accounts include a deficit in filtering noise from sensory channels, and a deficit in synchronizing theta-band oscillations in posterior temporal cortex with the 3–10 Hz rate of syllabic transition in speech.

Despite the lack of consensus with regard to neural models, there have been impressive findings with neonatal AEPs. A study by Molfese (2000), for example, recorded AEPs to consonant-vowel phonemes in 48 neonates within 36 hours of birth. Eight years later these children were evaluated with tests of reading and of IQ, and classified as *dyslexic*, *poor reader*, and *control*. Next, the now-8-year-old AEPs were averaged by group, and determined to differ by three components: the N1, the P2, and the N2. These AEPs yielded an astounding 81.3% correct classification of the sample! There are two basic science implications to highlight here. First, this illustrates, again, the importance of sensory/perceptual processing on high-level cognitive functions. Second, echoing what we saw with findings from Dehaene-Lambertz and colleagues, studying sensory processing in the very young infant can provide important insight about these systems. The practical implications? “If reading interventions were attempted shortly after birth on the basis of these data, 22 of 24 children in need of intervention at 8 years of age could have been targeted to receive intervention beginning at birth, while only 5 of 24 children who did not require intervention would have received it” (Molfese, 2000, pp. 241–242).

ALL TOGETHER NOW

Neurologist Robert Knight, neurosurgeon Nicholas Barbaro, and their colleagues have performed several electrocorticography (ECoG) studies in surgical epilepsy patients. *Web Link 18.3* illustrates, at an exquisite and astounding level of temporal and spatial precision, the dynamic interplay in “Lichtheim’s house,” and beyond, associated with a patient hearing a word and repeating it back to the experimenter (Edwards et al., 2009).

“AROUSING NOVEL IDEAS AND SUBTLE PERCEPTIONS AND JUDGMENTS” ... WITH THE WRITTEN WORD

Let’s conclude our whirlwind tour of the cognitive neuroscience of language with a brief peek at one element of the faculty of language that you have used each time you picked up this book – reading. More specifically, we’ll consider the real-world application of cognitive neuroscience research to understanding developmental disabilities with acquiring this skill (*Real-World Application 18.1*).

END-OF-CHAPTER QUESTIONS

1. Draw Lichtheim’s house model, and indicate how lesions of three different areas would produce three kinds of aphasia.
2. What level of language processing was of interest in the study of Dehaene-Lambertz et al. (2005) that varied place of articulation of stimuli? What about in the study of Abrams et al. (2013) that compared clips of speeches vs. spectrally rotated variants of these speeches?
3. What anatomical systems comprise the dual routes of language processing? What are their proposed functions?
4. What conclusions can be drawn from the “story” of FOXP2, with regard to genetic contributions to language function?
5. There are at least four distinct accounts of the functions of Broca’s area that we reviewed in this chapter. Summarize them and propose an account that you think best fits the current state of knowledge.
6. Explain the concept of predictive coding as it applies to sentence processing. What is the evidence for it?
7. Explain the proposed role of forward models in speech production, and the evidence for it.
8. What insights into reading have resulted from the study of dyslexia?

REFERENCES

- Abrams, Daniel A., Srikanth Ryali, Tianwen Chen, Evan Balaban, Daniel J. Levitin, and Vinod Menon. 2013. “Multivariate Activation and Connectivity Patterns Discriminate Speech Intelligibility in Wernicke’s, Broca’s, and Geschwind’s Areas.” *Cerebral Cortex* 23 (7): 1703–1714. doi: 10.1093/cercor/bhs165.
- Acheson, Daniel J., Massihullah Hamidi, Jeffrey R. Binder, and Bradley R. Postle. 2011. “A Common Neural Substrate for Language Production and Verbal Working Memory.” *Journal of Cognitive Neuroscience* 23 (6): 1358–1367. doi: 10.1162/jocn.2010.21519.
- Bates, Elizabeth, Stephen M. Wilson, Ayse Pinar Saygin, Frederic Dick, Martin I. Sereno, Robert T. Knight, and Nina F. Dronkers. 2003. “Voxel-Based Lesion-Symptom Mapping.” *Nature Neuroscience* 6 (5): 448–450. doi: 10.1038/nn1050.
- Bird, Helen, Matthew A. Lambon Ralph, Mark S. Seidenberg, James L. McClelland, and Karalyn Patterson. 2003. “Deficits in Phonology and Past-Tense Morphology: What’s the Connection?” *Journal of Memory and Language* 48 (3): 502–526. doi: 10.1016/S0749-596X(02)00538-7.

- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.
- Compston, Alastair. 2006. "On Aphasia. By L. Lichtheim, MD, Professor of Medicine in the University of Berne." *Brain* 1885; 7: 433–484." *Brain* 129 (6): 1347–1350. doi: 10.1093/brain/awl134.
- Dehaene-Lambertz, Ghislaine, Christophe Pallier, Willy Serniclaes, Liliane Sprenger-Charolles, Antoinette Jobert, and Stanislas Dehaene. 2005. "Neural Correlates of Switching from Auditory to Speech Perception." *NeuroImage* 24 (1): 21–33. doi: 10.1016/j.neuroimage.2004.09.039.
- Edwards, E., Srikantan S. Nagarajan, Sarang S. Dalal, Ryan T. Canolty, Heidi E. Kirsch, Nicholas M. Barabo, and Robert T. Knight. 2010. "Spatiotemporal Imaging of Cortical Activation during Verb Generation and Picture Naming." *NeuroImage* 50 (1): 291–301. doi: 10.1016/j.neuroimage.2009.12.035.
- Fedorenko, Evelina, John Duncan, and Nancy Kanwisher. 2012. "Language-Selective and Domain-General Regions Lie Side by Side within Broca's Area." *Current Biology* 22 (21): 2059–2062. doi: 10.1016/j.cub.2012.09.011.
- Fletcher, Paul. 1990. "Speech and Language Deficits." *Nature* 346 (6281): 226. doi: 10.1038/346226c0.
- Geschwind, Norman. 1970. "The Organization of Language and the Brain: Language Disorders after Brain Damage Help in Elucidating the Neural Basis of Verbal Behavior." *Science* 170 (3961): 940–944. doi: 10.1126/science.170.3961.940.
- Gopnik, Myrna. 1990a. "Feature-Blind Grammar and Dysphasia." *Nature* 344 (6268): 715. doi: 10.1038/344715a0.
- Gopnik, Myrna. 1990b. "Genetic Basis of Grammar Defect." *Nature* 347 (6288): 26. doi: 10.1038/347026a0.
- Graham, S.A., and S.E. Fisher. 2013. "Decoding the Genetics of Speech and Language." *Current Opinion in Neurobiology*. 23: 43–51. doi: 10.1016/j.conb.2012.11.006.
- Grodzinsky, Yosef, and Andrea Santi. 2008. "The Battle for Broca's Region." *Trends in Cognitive Sciences* 12 (12): 474–480. doi: 10.1016/j.tics.2008.09.001.
- Hagoort, Peter. 2005. "On Broca, Brain, and Binding: A New Framework." *Trends in Cognitive Sciences* 9 (9): 416–423. doi: 10.1016/j.tics.2005.07.004.
- Hickok, Gregory, and David Poeppel. 2000. "Towards a Functional Neuroanatomy of Speech Perception." *Trends in Cognitive Sciences* 4 (4): 131–138. doi: 10.1016/S1364-6613(00)01463-7.
- Houde, John F., Srikantan S. Nagarajan, Kensuke Sekihara, and Michael M. Merzenich. 2002. "Modulation of the Auditory Cortex during Speech: An MEG Study." *Journal of Cognitive Neuroscience* 14 (8): 1125–1138.
- Hurst, Jane A., Michael Baraitser, E. Auger, Frances Graham, and S. Norell. 1990. "An Extended Family with a Dominantly Inherited Speech Disorder." *Developmental Medicine and Child Neurology* 32 (4): 352–355. doi: 10.1111/j.1469-8749.1990.tb16948.
- Indefrey, Peter, and Willem J. M. Levelt. 2004. "The Spatial and Temporal Signatures of Word Production Components." *Cognition* 92 (1–2): 101–144. doi: 10.1016/j.cognition.2002.06.001.
- Kluender, Keith R., Randy L. Diehl, and Peter R. Killeen. 1987. "Japanese Quail Can Learn Phonetic Categories." *Science* 237 (4819): 1195–1197. doi: 10.1126/science.3629235.
- Kuhl, Patricia K., and James D. Miller. 1975. "Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants." *Science* 190 (4209): 69–72. doi: 10.1126/science.1166301.
- Kutas, Marta, and Kara D. Federmeier. 2011. "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)." *Annual Review of Psychology* 62: 621–647. doi: 10.1146/annurev.psych.093008.131123.
- Lai, Cecilia S. L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem, and Anthony P. Monaco. 2001. "A Forkhead-Domain Gene is Mutated in a Severe Speech and Language Disorder." *Nature* 413 (6855): 519–523. doi: 10.1038/35097076.
- Lieberman, Alvin M. 1996. *Speech: A Special Code*. Cambridge, MA: MIT Press.
- Makuuchi, Michiru, Jörg Bahlmann, Alfred Anwander, and Angela D. Friederici. 2009. "Segregating the Core Computational Faculty of Human Language from Working Memory." *Proceedings of the National Academy of Sciences of the United States of America* 106 (20): 8362–8367. doi: 10.1073/pnas.0810928106.
- McClelland, James L., and Karalyn Patterson. 2002a. "Rules or Connections in Past-Tense Inflections: What Does the Evidence Rule Out?" *Trends in Cognitive Sciences* 6 (11): 465–472. doi: 10.1016/S1364-6613(02)01993-9.
- McClelland, James L., and Karalyn Patterson. 2002b. "'Words or Rules' Cannot Exploit the Regularity in Exceptions." *Trends in Cognitive Sciences* 6 (11): 464–465. doi: 10.1016/S1364-6613(02)02012-0.
- Molfese, Dennis L. 2000. "Predicting Dyslexia at 8 Years of Age Using Neonatal Brain Responses." *Brain and Language* 72 (3): 238–245. doi: 10.1006/brln.2000.2287.
- Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.
- Pinker, Steven, and Michael T. Ullman. 2002a. "The Past and Future of the Past Tense." *Trends in Cognitive Sciences* 6 (11): 456–463. doi: 10.1016/S1364-6613(02)01990-3.

- Pinker, Steven, and Michael T. Ullman. 2002b. "Combination and Structure, Not Gradedness, Is the Issue." *Trends in Cognitive Sciences* 6 (11): 472–474. doi: 10.1016/S1364-6613(02)02013-2.
- Rogalsky, Corianne, and Gregory Hickok. 2011. "The Role of Broca's Area in Sentence Comprehension." *Journal of Cognitive Neuroscience* 23 (7): 1664–1680. doi: 10.1162/jocn.2010.21530.
- Tourville, Jason A., Kevein J. Reilly, and Frank H. Guenther. 2008. "Neural Mechanisms Underlying Auditory Feedback Control of Speech." *NeuroImage* 39 (3): 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054.
- van Berkum, Jos J. A., Peter Hagoort, and Colin M. Brown. 1999. "Semantic Integration in Sentences and Discourse: Evidence from the N400." *Journal of Cognitive Neuroscience* 11 (6): 657–671. doi: 10.1162/089892999563724.
- Vargha-Khadem, Faraneh, and Richard E. Passingham. 1990. "Speech and Language Deficits." *Nature* 346 (6281): 226. doi: 10.1038/346226c0.
- Vargha-Khadem, Faraneh, Kate Watkins, Katie Alcock, Paul Fletcher, and Richard E. Passingham. 1995. "Praxic and Nonverbal Cognitive Deficits in a Large Family with a Genetically Transmitted Speech and Language Disorder." *Proceedings of the National Academy of Sciences of the United States of America* 92 (3): 930–933. doi: 10.1073/pnas.92.3.930.
- Willems, Roel M., and Peter Hagoort. 2009. "Broca's Region: Battles Are Not Won by Ignoring Half of the Facts." *Trends in Cognitive Sciences* 13 (3): 101. doi: 10.1016/j.tics.2008.12.001.
- Young, Liane, Joan Albert Camprodon, Marc Hauser, Alvaro Pascual-Leone, and Rebecca Saxe. 2010. "Disruption of the Right Temporoparietal Junction with Transcranial Magnetic Stimulation Reduces the Role of Beliefs in Moral Judgments." *PNAS* 107 (15): 6753–6758. doi: 10.1073/pnas.0914826107.

OTHER SOURCES USED

- Boatman, Dana. 2002. "Neurobiological Bases of Auditory Speech Processing." In *The Handbook of Adult Language Disorders: Integrating Cognitive Neuropsychology, Neurology, and Rehabilitation*, edited by Argye E. Hillis, 269–291. New York: Psychology Press.
- Carroll, John B. 1958. "Communication Theory, Linguistics, and Psycholinguistics." *Review of Educational Research* 28 (2): 79–88. doi: 10.2307/1168862.
- Chomsky, Noam. 1959. Review of *Verbal Behavior*, by B.F. Skinner. *Language* 35 (1): 26–58. doi: 10.2307/411334.
- Fedorenko, Evelina, and Nancy Kanwisher. 2011. "Some Regions within Broca's Area Do Respond More Strongly to Sentences than to Linguistically Degraded Stimuli: A Comment on Rogalsky and Hickok." *Journal of Cognitive Neuroscience* 23 (10): 2632–2635. doi: 10.1162/jocn_a_00043.
- Gabrieli, John D. E. 2009. "Dyslexia: A New Synergy between Education and Cognitive Neuroscience." *Science* 325 (5938): 280–283. doi: 10.1126/science.1171999.
- Goswami, Usha. 2010. "A Temporal Sampling Framework for Developmental Dyslexia." *Trends in Cognitive Sciences* 15 (1): 3–10. doi: 10.1016/j.tics.2010.10.001.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298 (5598): 1569–1579. doi: 10.1126/science.298.5598.1569.
- Hickok, Gregory, and Corianne Rogalsky. 2011. "What Does Broca's Area Activation to Sentences Reflect." *Journal of Cognitive Neuroscience* 23 (10): 2629–2631. doi: 10.1162/jocn_a_00044.
- Pinker, Steven, and Ray Jackendoff. 2005. "The Faculty of Language: What's Special about It?" *Cognition* 95 (2): 201–236. doi: 10.1016/j.cognition.2004.08.004.
- Samet, Jerry. 1999. "History of Nativism." In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil, 586–588. Cambridge, MA: MIT Press.
- Serniclaes, Willy, Liliane Sprenger-Charolles, Rene Carre, and Jean-Francois Demonet. 2001. "Perceptual Discrimination of Speech Sounds in Developmental Dyslexia." *Journal of Speech, Language, and Hearing Research* 44 (2): 384–397. doi: 10.1044/1092-4388(2001/032).
- Skinner, B. F. 1957. *Verbal Behavior*. Action, MA: Copley.
- Wernicke, Carl. 1874. *Der Aphasische Symptomencomplex: Eine Psychologische Studie Auf Anatomischer Basis*. Breslau: Cohn and Weigert.

FURTHER READING

- Alexander, Michael P. 1997. "Aphasia: Clinical and Anatomic Issues." In *Behavioral Neurology and Neuropsychology*, edited by Todd E. Feinberg and Martha J. Farah, 133–149. New York: McGraw-Hill.
Authoritative overview of the eight classical aphasia syndromes.
- Binder, Jeffrey R. 2003. "Wernicke Aphasia: A Disorder of Central Language Processing." In *Neurological Foundations of Cognitive Neuroscience*, edited by Mark D'Esposito, 175–238. Cambridge, MA: MIT Press.
Exhaustive review of the neuropsychological and neuroimaging literature, which concludes that "Wernicke's aphasia, simply put, consists of [a] combination of ... syndromes and results from a larger lesion that encompasses both lexical-semantic and phoneme output systems" (p. 226).
- Chial, Heidi. 2008. "Mendelian Genetics: Patterns of Inheritance and Single-Gene Disorders." *Nature Education* 1 (1). <http://www.nature.com/scitable/topicpage/mendelian-genetics-patterns-of-inheritance-and-single-966>.
Primer on inheritance of single-gene disorders.
- Dehaene, Stanislas. 2009. *Reading in the Brain: The Science and Evolution of a Human Invention*. New York: Viking.
A thorough, reader-friendly introduction to the cognitive neuroscience of reading.
- Dronkers, Nina F. 2000. "The Gratuitous Relationship between Broca's Aphasia and Broca's Area." *Behavioral and Brain Sciences* 23 (1): 30–31. doi: 10.1017/S0140525X00322397.
A succinct argument, thoroughly referenced, against a grammar-specific view of Broca's area.
- Herholz, Sibylle C., and Robert J. Zatorre. 2012. "Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure." *Neuron* 76 (3): 486–502. doi: 10.1016/j.neuron.2012.10.011.
A comprehensive overview of the effects of musical training on the brain.
- Mahmoudzadeh, Mahdi, Ghislaine Dehaene-Lambertz, Marc Fournier, Guy Kongolo, Sabrina Goudjil, Jessica Dubois, ... Fabrice Wallois. 2013. "Syllabic Discrimination in Premature Human Infants prior to Complete Formation of Cortical Layers." *Proceedings of the National Academy of Sciences of the United States of America* 110 (12): 4846–4851. doi: 10.1073/pnas.1212220110.
Technically impressive demonstration of speech-like functions in the brains of prematurely delivered infants whose cortices are still dramatically underdeveloped.
- Salimpoor, Valorie N., Iris van den Bosch, Natasa Kovacevic, Anthony Randal McIntosh, Alain Dagher, and Robert J. Zatorre. 2013. "Interactions between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value." *Science* 340 (6129): 216–219. doi: 10.1126/science.1231059.
Integrates content from Chapter 14, Chapter 15, and Chapter 17 with "music appreciation": "aesthetic rewards arise from the interaction between mesolimbic reward circuitry and cortical networks involved in perceptual analysis and valuation."
- Scott, Sophie K., and Ingrid S. Johnsrude. 2003. "The Neuroanatomical and Functional Organization of Speech Perception." *Trends in Neurosciences* 26 (2): 100–107. doi: 10.1016/S0166-2236(02)00037-1.
Presents an alternative dual-route model of perception, with a greater emphasis on homologies with the monkey auditory system.
- Zatorre, Robert J., and Jackson T. Gandour. 2008. "Neural Specializations for Speech and Pitch: Moving beyond the Dichotomies." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 363 (1493): 1087–1104. doi: 10.1098/rstb.2007.2161.
A detailed consideration of what factors may underlie the lateralized processing of some sounds.

CHAPTER 19

CONSCIOUSNESS

KEY THEMES

- Consciousness can be construed at many levels, and the three that we'll consider here are physiological (*What physiological conditions are necessary to support consciousness?*), content-based (*What neural processes determine whether or not one is aware of a specific stimulus?*), and theoretical (*What are the neurobiological principles that make consciousness possible?*).
- The neurological syndrome of coma indicates that elevated activity in the brainstem reticular activating system is necessary for an individual to be conscious, but the persistent vegetative state/unresponsive wakefulness syndrome (UWS) indicates that elevated brainstem activity is not sufficient for consciousness.
- Lesions of the ventral pons can produce the diabolical locked-in syndrome (LIS), in which the patient is fully conscious, but only able to blink and produce vertical eye movements. LIS patients are often initially misdiagnosed as vegetative state/UWS.
- Perturbation of the brain with TMS, when coupled with EEG recording, reveals a systematic reduction of effective connectivity accompanying the loss of consciousness experienced during nonREM sleep, or while under general anesthesia.
- When visual stimuli are presented at the threshold of detection, V1 activity more closely tracks the subjective judgment of whether or not one has seen a target than the objective reality of whether or not a target was presented.
- In the monkey, detection of a target stimulus seems to depend not on the initial “feedforward sweep” of information from the retina and up through the visual hierarchy, but on reentrant, feedback signals to V1 after the feedforward sweep has been registered in higher-level visual areas.
- Studies using visual masking, as well as studies of the attentional blink, have been interpreted to suggest that widespread activation of high-level cortical areas corresponding to a putative “global workspace” might be necessary for awareness of a stimulus.
- One theory of consciousness, Integrated Information Theory, has given rise to an empirical measure, the Perturbational Complexity Index (PCI), that discriminates levels of consciousness associated with neurological syndromes, level of anesthesia, and stage of sleep.

CONTENTS

KEY THEMES

THE MOST COMPLEX OBJECT IN THE UNIVERSE

DIFFERENT APPROACHES TO THE PROBLEM

THE PHYSIOLOGY OF CONSCIOUSNESS

Neurological syndromes

Coma

Vegetative state/UWS

Recovery of consciousness from coma or UWS

Two dimensions along which consciousness can vary

Sleep

Anesthesia

Studying anesthesia with TMS/EEG

Studying anesthesia with surface EEG and deep brain EEG

Summary across physiological studies

BRAIN FUNCTIONS SUPPORTING CONSCIOUS PERCEPTION

Are we conscious of activity in early sensory cortex?

What intrinsic factors influence conscious visual perception?

Bottom-up and top-down influences on conscious visual perception

The role of reentrant signals to V1

Manipulating extrinsic factors to study conscious vs. unconscious vision

THEORIES OF CONSCIOUSNESS

Global Workspace Theory

Recurrent Processing Theory

Integrated Information Theory

A practical application of Integrated Information Theory

UPDATING THE CONSCIOUSNESS GRAPH END-OF-CHAPTER QUESTIONS

REFERENCES

OTHER SOURCES USED

FURTHER READING

THE MOST COMPLEX OBJECT IN THE UNIVERSE

Natural evolution, as brilliantly revealed by Charles Darwin (1809–1882), over these many million years gave rise to nervous systems as complex as the human brain, arguably the most complex object in the universe. And somehow, through the interactions among its 100 billion neurons, connected by trillions of synapses, emerges our conscious experience of the world and ourselves. (Laureys and Tononi, 2009, p. ix)

In 2005, to mark its 125th anniversary, *Science* magazine published a feature on “What don’t we know?”, posing 125 questions that science has yet to satisfactorily answer. Twenty-five of these were allocated a page in the journal and, although they weren’t explicitly rank-ordered, one of them necessarily had to appear on the first page of the section, and one on the 25th. Coming in at number two (and thereby reconfirming that “just because it appeared in *Science* doesn’t make it so”) was *What is the Biological Basis of Consciousness?* Number one had something to do with the origin of the universe. What?!? I mean, come on, without consciousness, the notion of what a universe is, not to mention that it can have an origin, wouldn’t exist ...

What is it that gives you that ineffable sense of knowing that you are who you are, that you are presently reading a cognitive neuroscience textbook, that you are (engrossed/tired/hungry/relaxed/stressed/whatever)? These are the questions at which we will take a crack in this, our final chapter together. AND IT’S GOING TO BE AWESOME!

DIFFERENT APPROACHES TO THE PROBLEM

Even though we all experience it, there’s an “unmeasurable” quality to consciousness that makes its study different from every other aspect of cognition that we have considered up to this point. There aren’t obvious ways in which behavioral measures like reaction time or selection of a response can index consciousness in the way that we can use these to study visual perception, or attention, or emotion, or sentence processing.

Nor are there a priori neural signatures that everyone would agree on as being indisputably related to consciousness, short of some basal level of physiological functioning in at least some parts of the brain. But even for these, as we’ll soon see, their invocation leads immediately to questions of the necessity vs. the sufficiency of activity in the systems in question. Nobody said that this would be simple.

The approach that we will take is to group cognitive neuroscience studies of consciousness into three categories: *physiological*, *content-based*, and *theoretical*. Physiological approaches to understanding consciousness will include quantitative measurement of metabolic or electrical activity of the brain during states of consciousness that can be independently validated, typically by a physician with specialized training. One area of study that has produced an enormous amount of knowledge about physiological bases of consciousness is coma and stages of recovery from coma. Two other models that are also yielding valuable information about relations between various physiological states and states of consciousness are sleep and anesthesia. Note that, for the “physiological/content-based” distinction to be useful, the former will be limited to physiological states that persist for minutes, hours, or longer. Shorter-lived, momentary states, such as the instantaneous phase of an oscillation in a particular frequency band in the EEG, will be grouped under the content-based approach to consciousness.

What we’ll call “content-based” approaches to consciousness are those that assess trial-by-trial awareness of specific stimuli or events, and seek to infer from trials with, versus trials without, awareness of the item in question, what are the neural correlates of awareness for that item/event. One such study that we have already considered was the phosphene induction study featured all the way back in *Research Spotlight 3.2*. Remember? The finding was that the likelihood that transcranial magnetic stimulation (TMS) of the occipital cortex would induce a phosphene was predicted by the power in the alpha band of the EEG during the 500ms prior to the TMS pulse. This tells us something about how the instantaneous state of V1 relates to visual awareness.

Finally, there are theoretical approaches, and here, we will limit ourselves to cognitive- and systems-level

neuroscience theories and models of how the brain might give rise to consciousness. Of course, it is invariably the case that tests of these theories will, themselves, be classifiable in our scheme as either “physiological” or “content-based” experiments, and so for some of this final section we’ll be considering how research that we had reviewed earlier fits into theory X as opposed to theory Y.

Alright, that’s the background, now let’s get into it.

THE PHYSIOLOGY OF CONSCIOUSNESS

Physiological states are objective, measurable phenomena, and so provide objective, measurable anchoring points to which we can remain tethered during our exploration of consciousness, and to which we can periodically return as things get progressively more abstract.

Neurological syndromes

From the syndromes of coma and the sequelae of minimally conscious state and locked-in syndrome, we can gain insight into which neuroanatomical and neurochemical systems are necessary for consciousness. From the vegetative state/unresponsive wakefulness syndrome (UWS; *Tip 19.1*), we can begin cataloguing physiological states that are not sufficient.

Tip 19.1

Due to the pejorative connotations of the term “vegetative” (despite its technical appropriateness), it has recently been proposed that the term “unresponsive wakefulness syndrome (UWS)” be adopted in the place of “vegetative state” (Laureys et al., 2010). Because much of the literature cited here predates 2010, however, there will necessarily be some switching back and forth.

Coma

Coma is a state of unarousable unconsciousness, characterized by a failure of the arousal/alerting system of the brain (ascending reticular activating system ...). For practical purposes, this includes failure of eye opening to stimulation, a motor response no better than simple withdrawal type movements and a verbal response no better than simple vocalization of nonword sounds. This presupposes that the motor pathways and systems that would allow the person to respond if he/she were conscious are intact. (Young, 2009, p. 138)

The “ascending reticular activating system.” We haven’t yet discussed this, have we? This is a name given to a collection of nuclei in the brainstem that are necessary for arousal from sleep and for maintenance of wakefulness. In particular, two nuclei located in the pons, the “bridge” (from Latin) of the brainstem, deliver the neurotransmitters acetylcholine (ACh) and glutamate to targets in the thalamus (*Tip 19.2*) via a dorsal pathway, and to the basal forebrain and posterior hypothalamus via a ventral pathway. Neurons in these structures, in turn, project diffusely throughout the brain, including cortex, releasing the excitatory/arousing neurochemicals glutamate, norepinephrine, histamine, and hypocretin/orexin. These pontine cholinergic nuclei, together with the cholinergic nuclei of the basal forebrain and the glutamatergic interlaminar nuclei of the thalamus, fire at high rates in wakefulness and during the rapid-eye-movement (REM) stage of sleep, but decrease or stop firing during nonREM stages of sleep (*Tip 19.3*). Thus, coma and related neurological states of compromised consciousness are often due to abnormal functioning of the brainstem (*Tip 19.4*). Due to the redundancy of nuclei and ascending fibers from the reticular activating system, recovery/transition from the initial state of coma often occurs within three weeks of coma onset.

Tip 19.2

The thalamic nuclei targeted by neurons of the reticular activating system are the interlaminar nuclei, sometimes referred to as “nonspecific thalamic nuclei,” because they seem to be involved in maintaining arousal, rather than processing any particular kind of information.

Tip 19.3

The reticular activating system has a reciprocal, antagonistic relationship with groups of GABAergic neurons located in the hypothalamus and basal forebrain. On a roughly 24-hour cycle, activity in these GABAergic neurons waxes and wanes, inhibiting the reticular activating system during sleep.

Tip 19.4

Causes of brainstem injury include (but are not limited to) stroke, metabolic and nutritional disorders, exogenous toxins, CNS infection and septic illness, seizures, hypothermia or hyperthermia, and frank structural damage (from Young, 2009).

At this point, we can provisionally say that elevated activity in the reticular activating system seems to be necessary for consciousness.

Vegetative state/UWS

The syndrome of vegetative state/UWS demonstrates, all too cruelly, that an activated state of the reticular activating system is not sufficient for consciousness. As summarized by Belgian neurologist Steven Laureys (2005):

Vegetative patients have their eyes wide open but are considered – by definition – to be unaware of themselves or their surroundings. They may grimace, cry or smile (albeit never contingent upon specific external stimuli) and move their eyes, head and limbs in a meaningless “automatic” manner. The vegetative state is often, but not always, chronic (the “persistent vegetative state”). Given proper medical care (i.e. artificial hydration and nutrition) patients can survive for many years. (p. 556)

Vegetative state/UWS is the state into which patients transition from coma. How is it determined that these patients lack consciousness? The absence of consciousness is inferred from the absence of any of the following signs across repeated neurological examinations that are ideally performed at different times of the day – no meaningful response to external

stimuli, including threats and noxious (i.e., painful) stimuli, no control over bladder or bowels, no sustained visual-pursuit eye movements, nor volitional fixation, no purposeful limb withdrawal, or localization of a stimulus with a limb. By the classical terminology, the diagnosis transitions to persistent vegetative state if present for longer than one month.

Because the transition out of coma necessarily entails a recovery of at least some of the functions of the reticular activating system, UWS is often accompanied by a daily cycle of opening and closing of the eyes. Even during periods of “wakefulness,” however, global brain metabolic activity in UWS rarely surpasses 50% of the level that is typical for the healthy awake state. For comparison, this 50%-of-normal level is similar to what is observed when healthy individuals are under general anesthesia or deep states of nonREM sleep.

The diagnosis of vegetative state/UWS inescapably has a dissatisfying, equivocal element because, short of brain-death, one can never know with absolute certainty that there isn’t some residual, difficult-to-detect spark of awareness “somewhere in there.” Indeed, the reality of the neurological syndromes of minimally conscious state and locked-in syndrome, the focus of *Real-World Applications 19.1*, play into this. For me, it’s analogous to (although, of course, much more consequential than) a failure to reject the null hypothesis in classical hypothesis testing – because one can never “prove” the null, one can never know with certainty that a “true” difference isn’t lurking somewhere in those data, but just too subtle to detect. And if that weren’t enough, there’s the work of indefatigable cognitive neuroscientist Adrian Owen and colleagues, whose fMRI scanning of diagnosed UWS patients is reviewed in *Research Spotlight 19.1*.

As implied in *Real-World Applications 19.1*, the classic LIS results from damage to the ventral brainstem, meaning that it often spares fibers carrying incoming sensory signals, including ascending somatosensory fibers. Indeed, many LIS patients report touch and pain sensations. These sensory fibers are also often spared in UWS patients, from whom “early” auditory and somatosensory ERP components can be recorded, as well as stimulus-evoked fMRI activity in primary sensory cortex. In UWS, however, these fMRI-measured responses typically do not propagate to higher-level cortex, as they do in neurologically healthy subjects.

REAL-WORLD APPLICATION**19.1 “What is it like to be conscious, but paralyzed and voiceless?”**

Horrifying, right? In my pre-science days, ideas that pushed me toward this career included the musing that one can lose an arm, or a leg, or both, and still be the same person. Lose just a tiny piece of brain from the wrong spot, however, or tweak its function with a pharmacological agent, and one's very essence could be dramatically altered. (And this was long before I had encountered the expression “He was no longer Gage”...) What if one were to take this hypothetical exercise in a different direction, losing the use of *all* of one's body, but leaving the brain intact? Unfortunately, this is not just an idle, lay-philosophical question for patients with locked-in syndrome (LIS).

LIS typically results from damage to the corticospinal and corticobulbar pathways, most often due to stroke at the level of the ventral pons. The former pathway we learned about in *Chapter 7*, the latter carries fibers from the motor cortex to motor neurons controlling the muscles of the face, head, and neck (i.e., the lower cranial nerves). An authoritative definition of LIS is:

a state in which selective supranuclear motor de-efferentation produces paralysis of all four limbs and the last cranial nerves without interfering with consciousness. The voluntary motor paralysis prevents the subjects from communicating by word or body movement. Usually, but not always, the anatomy of the responsible lesion in the brainstem is such that locked-in patients are left with the capacity to use vertical eye movements and blinking to communicate their awareness of internal and external stimuli. (Plum and Posner, 1983, p. 9)

Despite the sparing of vertical eye movements and blinking in most cases of LIS, upon transition from coma, many patients with LIS go misdiagnosed as UWS for extended periods of time. One celebrated case, the Frenchman Jean-Dominique Bauby, who dictated a bestselling memoir after his stroke (1997), also founded the Association for LIS (ALIS). Laureys et al. (2005) cite a survey of 44 patient-members of ALIS, which indicated that the average time between insult and diagnosis of LIS was 78 days, and for some, more than four years! Just imagine being the patient who has written that the doctors believed he was a “vegetable and treated him as such” (Laureys et al., 2005, p. 500). The same ALIS survey indicated that family members made the initial observation that their loved-one might be conscious in 55% of the cases, doctors doing so in only 23% of cases. Therapists are also often credited with making the discovery. In one case a therapist was testing the gag reflex of a presumed vegetative patient, and the patient bit the therapist. As the therapist cursed and withdrew his finger, the patient managed a grin. The now-suspicious therapist asked, “How much is $2 + 2$?” and the patient blinked four times.

The eloquent writing of many patients with LIS leaves no doubt about their level of consciousness. Bauby (1997) dictated his book letter-by-letter, blinking each time his therapist spoke the desired letter from a frequency-ordered list. Phillippe Vigand (2000), the finger biter, composed his book on software controlled by eye movements tracked with an infrared camera.

RESEARCH SPOTLIGHT

19.1 “Detecting awareness in the vegetative state” (Owen et al., 2006)

Awareness in the vegetative state? How could this be? It has already been stated, *in this book*, that “vegetative patients ... are considered – by definition – to be unaware of themselves or their surroundings.” Answer: it could be so if the definition is wrong, or if the diagnosis is wrong (in which case the title might better have been “Detecting awareness in patients previously diagnosed as being in a vegetative state,” huh?). Let’s see.

One of the first in the series of fMRI studies by Owen and colleagues – which has perhaps garnered more attention in the popular press than any other cognitive neuroscience research program ever – was an fMRI study of a 23-year-old traumatic brain injury patient who had been in a persistent vegetative state for five months. First, it was observed that for spoken sentences vs. acoustically matched noise sequences (à la Abrams et al., 2013, from *Chapter 18*), activity in posterior STG and MTG was greater for speech vs. noise. A tantalizing result, but not definitive, as it could possibly be attributable to automatic activation of auditory/language cortex spared by the accident. The second task was devised to require a volitional, internally generated response: the patient was asked, at specific times during the scan, to either “imagine yourself playing tennis,” or “imagine yourself walking around your house, visiting each room.” The results, illustrated in *Figure 19.1*, indicated that the patient’s neural activity in response to these two sets of instructions was comparable to that of control subjects.

Now, of course, this is an analytically simplistic example of “brain reading” relative to the sophisticated MVPA methods that we’ve discussed in previous chapters. Indeed, had this patient been scanned absent any comparison, it would be inferentially problematic to visually inspect activity from the scan and conclude that she was imagining tennis vs. navigating her house. However, that’s not the point here. The point is that

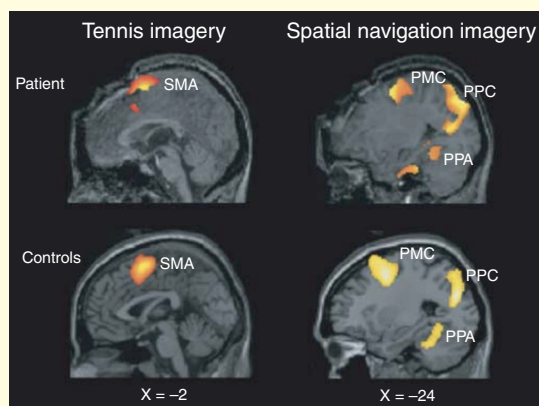


FIGURE 19.1 fMRI responses to imagery instructions, in a patient diagnosed as being in a persistent vegetative state (top row) and in the group-average of 12 control subjects. Source: Owen, Coleman, Boly, Davis, Laureys, and Pickard, 2006. Reproduced with permission of AAAS.

Research Spotlight 19.1 *Continued*

this study indicated that this patient, who was believed to be vegetative, was, in fact, able to modulate her brain activity “on command,” in more or less the same way as did unequivocally conscious control subjects. Thus, one might reasonably conclude that she, too, was conscious, despite also being fully locked in, such that she was deprived of even volitional eye movements or blinks.

In a subsequent study, Monti et al. (2010) scanned 54 patients with disorders of consciousness, and found five who could willfully modulate their brain activity in this way. With one patient they took the further step of telling him that he’d be asked a series of autobiographical questions, and that he was to respond “yes” by imagining tennis, or “no” by imagining navigating through his house. In response to the question “Is your father’s name Alexander?,” for example, he correctly responded “yes” by producing elevated activity in pre-defined “tennis regions.” To the question “Is your father’s name Thomas?,” he correctly responded “no” by producing elevated activity in navigation regions.

And so, which is it? A “wrong” definition or “wrong” diagnoses? The answer seems to be that these patients, to return to my earlier analogy to hypothesis testing, are cases of “Type II error” – bedside diagnostics for them lacked the sensitivity to rule out the default diagnosis of UWS. (More on Owen’s research at [Web Link 19.1.](#))

Recovery of consciousness from coma or UWS

In some fortunate instances, patients in a coma or UWS experience substantial recovery of function. Such cases offer unique opportunities to assess the changes in the brain that accompany recovery of conscious awareness. Of course, doing so requires having thoroughly documented the patient’s condition prior to recovery, something that is not standard procedure at most hospitals. One notable exception is the Coma Science Group, under the direction of Steven Laureys, at the Université de Liège in Belgium, which has produced a prodigious amount of groundbreaking research on the neurology of consciousness. Their work indicates that, although the UWS is associated with widespread metabolic dysfunction, measures of global metabolic activity, alone, lack both the sensitivity and the specificity that one would want in a “consciousness meter”: there are cases of patients who have recovered from UWS without significant changes in brain metabolism; and there are neurologically healthy volunteers whose global brain metabolism levels are as low as those of UWS patients.

At least one intriguing case suggests that more diagnostic precision might be found in patterns of connectivity than in measures of raw metabolic activity.

A patient was scanned with $H_2^{15}O$ PET two weeks after the onset of UWS, and again, at rest, four months after recovery of consciousness and “partial autonomy.” While in UWS, functional connectivity between the thalamus and several regions of lateral PFC (BAs 8, 9, and 10) and anterior cingulate cortex (BAs 24 and 32) was reduced relative to 18 control subjects; after clinical recovery, thalamocortical connectivity was also recovered to normal levels. Further consistent with the disconnection idea is that UWS patients (who, unfortunately, have not recovered) show “‘Functional disconnections’ in long-range cortico-cortical (between latero-frontal and midline-posterior areas) and cortico-thalamo-cortical (between non-specific thalamic nuclei and lateral and medial frontal cortices) pathways” (Laureys, 2005, p. 557).

Two dimensions along which consciousness can vary

This review of neurological syndromes has introduced the idea of two dissociable dimensions along which states of consciousness can vary. One, which seems to be necessary, but is clearly not sufficient, is the level of physiological arousal of the brain. This is portrayed as the horizontal axis in the diagram in *Figure 19.2*. The

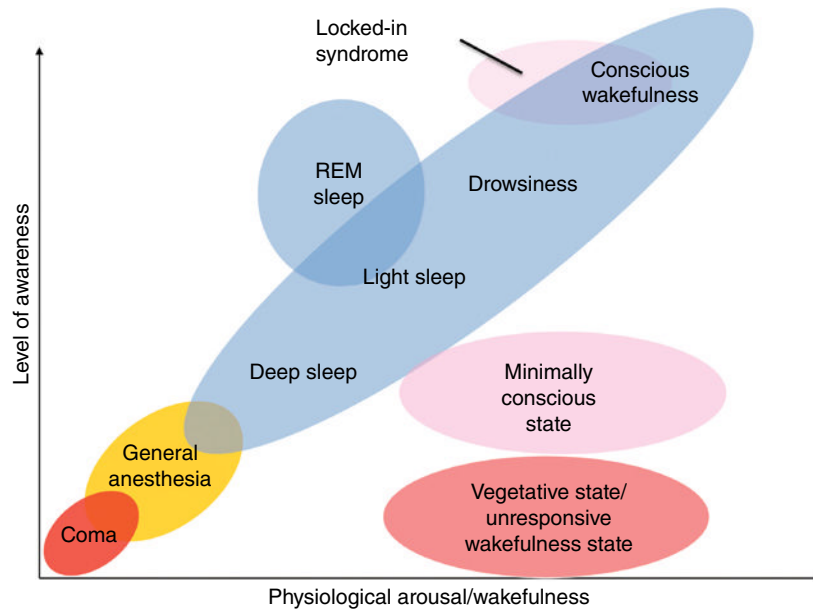


FIGURE 19.2 A depiction of neurophysiological states as they fall along two dimensions that are related to consciousness: arousal (horizontal axis) and awareness (vertical axis). The blue region corresponds to normal physiological states, pink to pathological coma and orange to pharmacologically induced coma (a state from which patients cannot be awakened). Source: Adapted by the author from Laureys, 2005. Reproduced with permission of Elsevier.

second, for the time being, we can label the “level of awareness.” Different clinical states that we just discussed, and physiological states that we’re about to take up in the next two subsections, can be portrayed in terms of where they fall in this two-dimensional space. The label, and the idea behind, the second dimension is not satisfactory, because it’s just a descriptive term. Ideally, to really understand the neural bases of consciousness, one would want to know what measurable, physical property varies along that axis, “bringing awareness along with it.” Indeed, one way to construe the third section of this chapter, on theoretical approaches to consciousness, will be to evaluate each theory in terms of how well its explanatory factor would do if plugged into the label for the vertical axis of this plot. (To make this concrete, the implication of one theory will be that the physical property that varies along the vertical axis is connectivity – greater levels of connectivity within the brain, coupled with greater levels of physiological arousal, result in higher levels of awareness.)

In the next two subsections, on sleep and anesthesia, we will introduce a different technique for relating physiological state to consciousness: estimating effective connectivity from the TMS-evoked response in the EEG. Once we have established an understanding of this technique, the final subsection of this section on “physiological state” will return to coma, and summarize how the recovery of consciousness is paralleled by a recovery of large-scale connectivity within the brain.

Sleep

Sleep offers another fascinating window on consciousness. As discussed previously, sleep results from an inhibition of arousing signals to thalamus and cortex from the brainstem and basal forebrain. In the absence of these signals, two phenomena come about. First, there is a marked reduction in long-range connectivity between brain areas. Second, there is a marked increase in local homogeneity of activity

across populations of cortical neurons whose waking activity is typically not correlated, certainly not for extended periods of time. At the level of the EEG, the first change to occur at the transition between quiet eyes-closed wakefulness and sleep is the replacement of the prominent “high-voltage” alpha rhythm (*Figure 3.10*) with low-voltage oscillations in the theta band. Then, the EEG gradually comes to be dominated by the large-amplitude low-frequency ($\sim 1\text{--}2\text{ Hz}$) “slow waves” (a.k.a. delta waves) of so-called “slow-wave sleep.” High-density EEG recordings have revealed slow waves to be traveling waves that sweep across the cortical surface, typically from anterior to posterior, in a manner very reminiscent of the expanding, concentric rings produced by a single stone dropped into the proverbial flat pond: an expanding arc of neuronal depolarization, and, hence, “lock-step” elevated firing, followed by a prolonged, few hundred milliseconds period of hyperpolarization, during which synaptic activity is reduced to almost nil. Upon falling asleep, one typically progresses through three stages of ever-deeper nonREM sleep, with only infrequent slow waves during stage N1, roughly five per minute in N2, and more than 10 per minute in N3. After a bout of several, if not tens of, minutes in N3 (longer at the beginning of the night, progressively shorter as the night wears on), one transitions back into N2, then “pops” into REM.

REM sleep shares many characteristics with wakefulness, in that it entails high activity in the reticular activating system, basal forebrain, and thalamus. These, in turn, arouse the cortex via release of ACh and glutamate, just as in wakefulness. The difference between REM sleep and wakefulness is that additional neurotransmitter systems also driven by the reticular activating system during wakefulness – including norepinephrine, histamine, hypocretin, and serotonin – remain inhibited during REM. Additionally, a set of neurons in the dorsal pons applies tonic inhibition of muscle tone.

REM is, of course, the stage of sleep during which we experience our most vivid dreams. Dreams can be construed as “sleep consciousness,” and offer yet another window through which consciousness can be studied (see *Further Reading* by Nir and Tononi, 2010).

Particularly germane for our consideration of physiological states associated with consciousness are

the systematic changes in the TMS-evoked response in the EEG as individuals transition from quiet wakefulness into sleep, as described in *Methodology Box 19.1*, and *Web Clip 19.1*.

Anesthesia

[M]any aspects of an EEG of anaesthesia are really consistent with patients who are in a coma. The key difference is that anaesthesia is reversible ... If you came in for an operation and I told you I was going to put you into a coma, you would probably get up and run away! [But] the idea of being put to sleep is a euphemism ... For your surgery you need to be in a state of general anaesthesia, not sleep. (Anesthesiologist and neuroscientist Emery Brown, quoted in Heathcoat, 2011)

Anesthesia refers to a chemically induced, reversible, loss of sensation, and general anesthesia, of interest to us here, to a chemically induced loss of consciousness. As detailed in the *Further Reading* by Brown, Lydic, and Schiff (2010), and illustrated in *Figure 19.2*, general anesthesia is, indeed, a physiological state more akin to coma than to sleep. One can be shaken awake from even the deepest sleep; not so with anesthesia (or coma). Further, with one exception (*Tip 19.5*), anesthesia causes a decline in overall brain glucose metabolism, as assessed with PET. The effects of anesthesia are dose dependent. For example, the anesthesiologist Michael Alkire describes a patient for whom a small dose of desflurane lowered global cerebral metabolism by 5–10% from baseline, rendering the patient sedated, but still able to generate a movement on request. The patient lost consciousness at a larger dose, which reduced global cerebral metabolism by 30%.

There are two additional findings from anesthesia that we will consider here, the first relating to a “breakdown of effective connectivity” reminiscent of what is seen with sleep. The second, neuroimaging evidence for why this might happen.

Studying anesthesia with TMS/EEG

A second study by Tononi’s group, led by Fabio Ferrarelli and including anesthesiologist Robert Pearce, replicated the general procedures of the study featured in *Methodology Box 19.1*, but did so while subjects were administered the anesthetic agent midazolam.

METHODOLOGY BOX

19.1 Studying sleep (and consciousness) with simultaneous TMS-EEG

Giulio Tononi, trained as a psychiatrist and a neuroscientist in Italy, and now based at this author's institution in Madison, WI, has been at the forefront of using TMS in innovative ways to address questions of basic neuroscience, including sleep, anesthesia, and consciousness. In 2005, a team led by then-fellow Marcello Massimini laid out the following rationale:

Using TMS/EEG to investigate critical differences in the functioning of the waking and sleeping brain offers several advantages. Unlike sensory stimulation, direct cortical stimulation does not activate the reticular formation and bypasses the thalamic gate. Thus, it directly probes the ability of cortical areas to interact, unconfounded by peripheral effects. (Massimini et al., 2005, p. 2229)

The experiments were run late at night, when subjects would more readily fall asleep. Subjects were fitted with an electrode cap connected to an EEG amplifier that, when triggered by the TMS stimulator, would momentarily divert the signal coming from the electrodes, such that the amplifier would not saturate in response to the TMS pulse. (The voltage transient produced by a pulse of TMS exceeds the range within which a conventional EEG amplifier, configured to detect exquisitely small voltage fluctuations on the scalp, can detect signals. The resulting “saturation” would render the amplifier insensitive to voltage fluctuations for hundreds of milliseconds or longer, meaning that the response evoked by the TMS pulse would not be recorded.)

Figure 19.3 illustrates the EEG response to a TMS pulse delivered to the right premotor cortex of a single subject before falling asleep (A–C) and during nonREM sleep (A'–C'). A and A' show the traces from all 60 electrodes in the cap, all plotted on the same timeline, and color-coded red during periods when the signal deviated significantly from baseline. Comparison across the two conditions illustrates that (1) the initial TMS-evoked response was roughly twofold larger in magnitude during nonREM sleep than during wakefulness, and (2) the TMS-evoked response was of considerably longer duration during wakefulness than during nonREM sleep. The reason for the latter is evident from C and C': whereas in wakefulness the initial response of the region directly under the coil was followed by responses in contralateral and posterior regions, then again in right PMC and then again in the left; during nonREM sleep the response stayed local. Thus, during wakefulness the ERP was “kept alive” longer because activity in proximal areas triggered activity in distal areas, which then re-activated the proximal area, which then re-activated the distal area. There was greater effective connectivity during wakefulness. (Recall that *functional connectivity* refers to correlated activity between two regions; *effective connectivity* to one region exerting a causal influence over another.)

Why was the initial response of such greater magnitude during nonREM sleep? One possibility is that, because proximal neurons are more tightly synchronized during nonREM sleep, the response to TMS in the region that was directly stimulated was also more synchronized. A second is that thalamic neurons are more likely to respond in burst mode (*Figure 3.7*) during nonREM sleep, because they are in a hyperpolarized state.

The point emphasized by the authors is that the fading of consciousness as we fall asleep is accompanied by a breakdown of long-range connectivity.

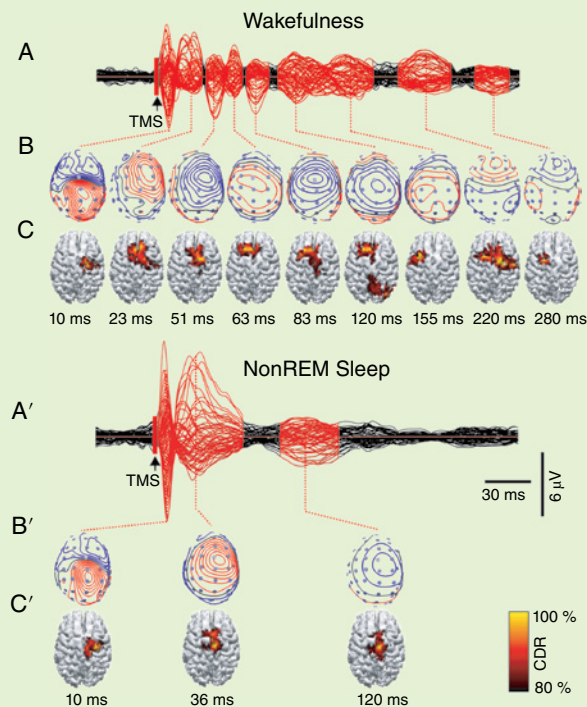
19.1 *Continued*

FIGURE 19.3 TMS-EEG from a single subject during wakefulness and nonREM sleep. Rows A and A' illustrate so-called “butterfly plots” in which the trace from each electrode is plotted along the same time axis. B and B' show the scalp distribution of voltages associated with the peak of each significant excursion of the signal. C and C' show the source solutions of the maps from B and B', displayed on the subject's MRI. CDR=current density reconstruction. Source: Massimini, Farrarelli, Huber, Esser, Singh, and Tononi, 2005. Reproduced with permission of AAAS.

Tip 19.5

The anesthetic agent ketamine is “dissociative,” in that it lowers awareness, and affects different brain regions in different ways, but, overall, causes an increase in global brain metabolic activity. Further, not unlike the dissociated neurological states that we have considered, including UWS, the eyes remain open during ketamine anesthesia.

Midazolam is a benzodiazepine whose anesthetic properties derive from activating GABA_A receptors, and thus imposing widespread inhibition in the brain. The effects of midazolam on the TMS-evoked response were virtually identical to those seen with nonREM sleep: the immediate response under the TMS coil was of greater magnitude, but the overall response was markedly reduced in terms of anatomical spread and in terms of temporal duration. Thus, anesthesia-induced loss of consciousness is also associated with drastic reduction of effective connectivity.

Studying anesthesia with surface EEG and deep brain EEG

Despite the considerable strength of the TMS/EEG approach that we have described, it is limited, as are all EEG studies, by the fact that it can't directly "see inside" the brain. This problem has been addressed by a team based in Marseille, France, led by anesthesiologist Lionel Velly, that has studied the effects of general anesthesia on patients with Parkinson's disease (PD) who have electrodes chronically implanted in the subthalamic nucleus of the basal ganglia (see *Figure 7.10*). Signal from the in-dwelling electrodes provided subcortical field potential recordings, referred to as ESCoG, that can be thought of as a "subcortical EEG." They performed two analytically independent analyses, one being a fairly conventional spectral analysis of the electrical signals, a second being a nonlinear measure based on deterministic chaos theory. "Schematically," as Velly et al. (2007) explain it, dimensional activation (D_a) estimates "the complexity of EEG and ESCoG signals: The more alert the patient is, the more complex the EEG is, and the higher the D_a is; the deeper the sleep or anesthetic state is, the simpler the EEG is, and the lower the D_a is" (pp. 204–205).

The findings from this study, illustrated in *Figure 19.4*, are striking and clear: anesthesia-induced loss of

consciousness is associated with a precipitous drop in oscillatory power at frequencies above the theta range, and a comparably precipitous drop in D_a in the cortical EEG. Measures of subcortical activity, in contrast, begin a gradual decline with anesthesia onset. This suggests that anesthesia-related loss of consciousness, and perhaps also the commensurate loss of effective connectivity reported in the TMS/EEG study of Ferrarelli et al. (2010), is primarily due to effects in cortex.

Summary across physiological studies

Let's take a moment to consolidate what we've learned about the "physiology of consciousness." First, a factor that is necessary, but not sufficient, for conscious awareness is elevated activity in the reticular activating system, which, in turn, drives the state of arousal in thalamus and cortex. When this system is in an inactive state, whether due to neurological dysfunction (e.g., coma), sleep state, or anesthesia, one does not have conscious awareness. That elevated levels of physiological arousal are not sufficient to support consciousness, however, is seen in vegetative state/UWS, as well as in circumstances that we have not discussed, sleepwalking and epileptic seizures. What do UWS patients, and patients under anesthesia, lack that

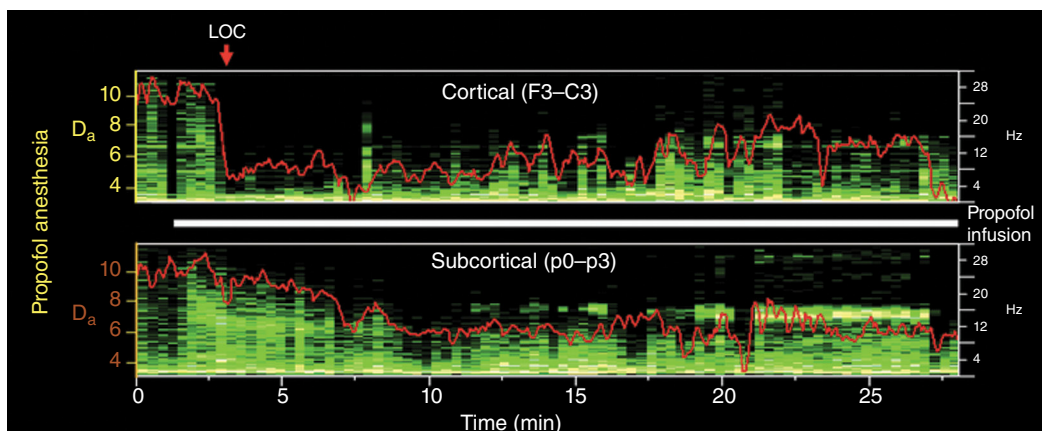


FIGURE 19.4 The time course of propofol anesthesia-induced loss of consciousness ("LOC"), as indexed by electrical activity measured from scalp EEG ("Cortical (F3–C3)") and from electrodes implanted in the subthalamic nuclei ("Subcortical (p0–p3)"). Red traces indicate level of dimensional activation (D_a) computed from the electrical signals (see text), with D_a axis on left-hand side of graphs, and green-to-white horizontal bars indicate spectral power in the EEG in 20 second time steps, with axis for the spectrograms on the right-hand side of the graphs. Source: Velly, Rey, Bruder, Gouvitsos, Witjas, Regis, Peragut, and Gouin, 2007. Reproduced with permission of Wolters Kluwer.

individuals who are conscious (REM sleep, LIS, healthy wakefulness) seem to have? Complex cortical activity and substantial cortico-cortical and cortico-thalamic connectivity. In the final section of this chapter, we will consider some theoretical explanations for why these aspects of brain activity might be important for consciousness.

As a “farewell to physiology,” *Figure 19.5* illustrates the evolution of the TMS/EEG response as subjects recover from coma, indicating that, just as with sleep and with anesthesia, a return to consciousness in these patients is accompanied by an increase in complex, interconnected activity.

FIGURE 19.5 Tracking recovery from coma with TMS/EEG. Source: Rosanova, Gosseries, Casarotto, Boly, Casali, Bruno, Mariotti, Boveroux, Tononi, Laureys, and Massimini, 2012. Reproduced with permission of Oxford University Press.

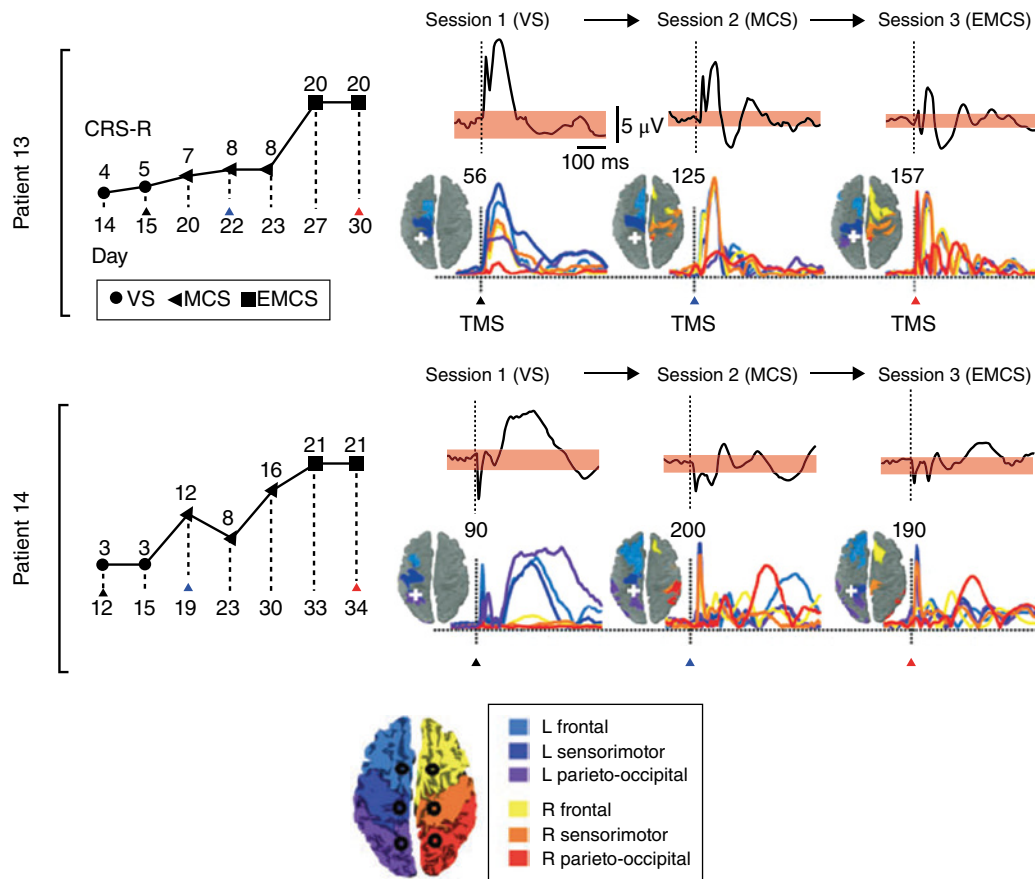


FIGURE 19.5.A Two patients whose recovery transitioned from coma to vegetative state (VS) to minimally conscious state (MCS) to emerged-from-MCS (EMCS), at which point they had recovered the ability to communicate and were classified as “recovered” by the Coma Recovery Scale-Revised (CRS-R), a standard diagnostic assessment. CRS-R score is tracked as a function of day post-trauma, with arrows indicating day of each TMS/EEG session (black = first; blue = second; red = third). At the top of the diagram for each session is the TMS-evoked response from the electrode under the coil. Lower left illustrates TMS target (white “+”) and brain regions from which a TMS-evoked response occurred, color-coded as per the legend. Plots to the right of each brain show the estimated TMS-evoked current from each of the six sources identified with black circles in the legend. The numbers between the two plots (e.g., 56, 125, 157 for Patient 13) indicate the number of vertices in the cortical mesh from the source solution at which a significant TMS-evoked response was detected.

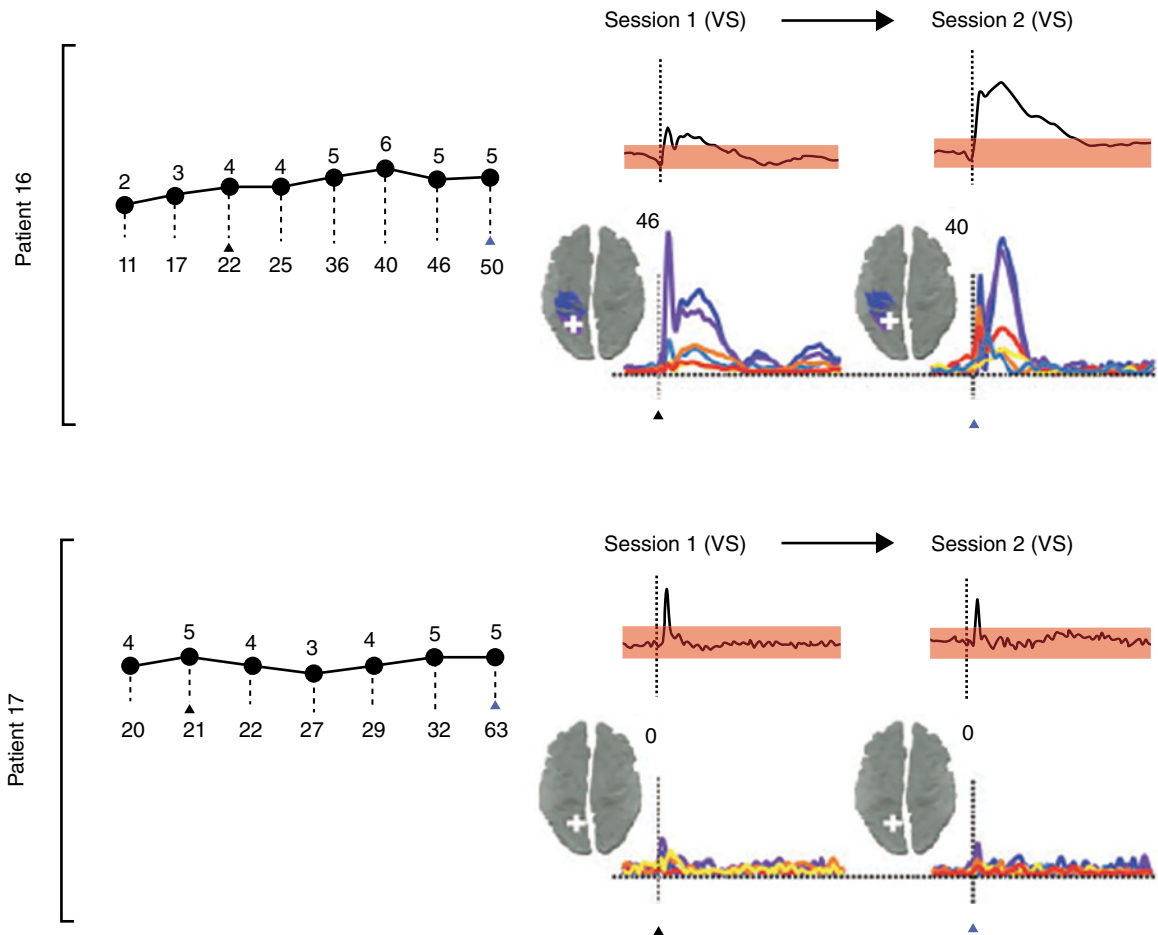


FIGURE 19.5.B Data from two patients who did not emerge from VS.

BRAIN FUNCTIONS SUPPORTING CONSCIOUS PERCEPTION

Up until now we've considered what factors underlie the state of being conscious. Now we turn to a different question: *What determines the contents of our conscious awareness at any given moment?* As I lift my eyes from my computer screen and look out from my front patio, what neural functions underlie my becoming aware of that sign across the street, or the hawk that's perched up in the tree? It's presumably not "just" the visual processes that we considered earlier in this book, because, as we saw with the phenomenon of the attentional blink (*Methodology Box 17.1*), one can

be presented with the same stimulus under identical external conditions and sometimes one becomes aware of it and other times one doesn't. And so minimally, we would need to specify what is different about the state of the visual system during trials when we are aware of the stimulus vs. during trials when we are not. And then there's the question of attention. Patients with hemispatial neglect are not aware of stimuli on the left side of space, so surely attention and consciousness are closely linked? Undoubtedly they are, but one of the claims that we'll consider in this section is that the two can be dissociated.

Research that addresses these kinds of questions is often referred to as the study of the neural correlates

of consciousness (NCC). As alluded to in the previous paragraph, studies of the NCC often use the strategy of presenting subjects with a series of identical trials, under conditions when they may or may not be aware of the stimulus. After the experiment, trials are sorted as a function of whether the target was or was not detected, and trials from the two types are compared for differences. Although we will restrict ourselves here to the visual modality, we'll see that there are many domains of cognition that are implicated in this research.

Are we conscious of activity in early sensory cortex?

The first question we might want to address when thinking about visual awareness is: what level(s) of the system potentially *can* support it? Let's say, for example, we had some way of determining that areas V1–V4 in the ventral stream are very machine-like, in that although the output of their processing is raw material that is needed for visual awareness, they don't actually contribute to the generation of visual awareness. If we knew this to be true, then we could focus our attention on higher level visual areas and not worry about fluctuations in signal in the earlier areas – these might relate to something like signal strength, but couldn't tell us about awareness, per se.

There are two classes of experimental procedure that have been used extensively to address this question. The one that we will consider in detail is, in a way, the simplest: it consists in presenting a single item on each trial, and comparing brain activity from trials in which the stimulus was seen vs. from trials in which it wasn't. A second experimental strategy for addressing this question takes advantage of the phenomenon of **binocular rivalry**: when a different image is presented to each of the two eyes, the percept is of seeing just one of the two, with frequent seemingly spontaneous “switches” between which of the two one is currently seeing. Because it explicitly dissociates the information being processed by the retina from the percept that we are consciously aware of, this procedure has been used extensively in studies of the NCC in both humans and monkeys. It's a rich and nuanced literature to which we can't do justice here, but that can be explored via publications by Leopold and Logothetis

(1996), Crick (1996), Engel et al. (1999), and Blake and Tong (2008), all listed under *Further Reading*.

Now, back to “single-item” studies. The first that we will consider in some detail is a very simple, very elegant fMRI study by cognitive neuroscientists David Ress and David Heeger (2003). It was designed to ask, in effect, *Does activity in early visual cortex correspond to the external stimulus from which photons are impacting our retinæ, or does it correspond to our internal perception?* Subjects fixated centrally, and every two seconds a ring of “background” visual noise would appear for one second, and on a randomly determined 17% of trials a target (a sinusoidal grating; i.e., parallel lines) would be superimposed on the background (Figure 19.6.A). Subjects responded “yes” or “no” on every trial, and target contrast was adjusted for each subject such that s/he would detect it on roughly 80% of trials. Each trial could have one of four outcomes: the target is present and the subject detects it (a “hit”); the target is present and the subject does not detect it (a “miss”); the target is absent and the subject reports detecting it (a “false alarm”); or the target is absent and the subject does not detect it (a “correct rejection”). It was assumed that hits would produce a robust stimulus-evoked response in V1, and correct rejections were treated as the baseline. The questions of interest, then, were how V1 would respond on trials in which the physical display and the subject's percept were at odds. If the fMRI response to misses resembled that of hits, and the fMRI response to false alarms

FIGURE 19.6 Does V1 activity correspond to the external world, or to our conscious perception of it? Source: Ress and Heeger, 2003. Reproduced with permission of Nature Publishing Group.

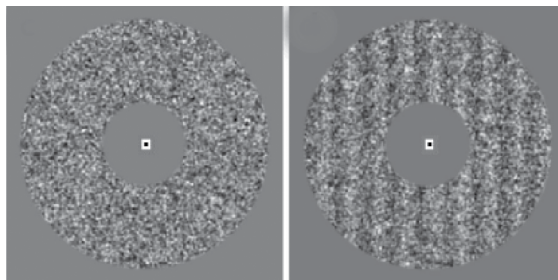


FIGURE 19.6.A A stimulus without the target vertical grating (left), and with it (right).

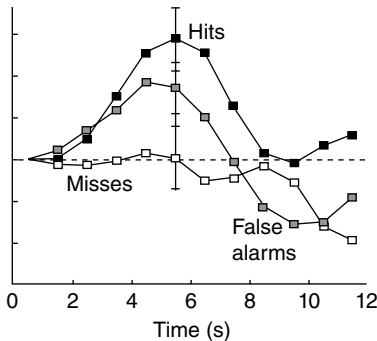


FIGURE 19.6.B fMRI responses from V1 of a single subject, time-locked to stimulus onset, as a function of behavioral response. Correct rejection trials were treated as the baseline for this analysis, meaning that the activity during misses did not differ from activity during correct rejections. (Each tick mark on the vertical axis corresponds to 0.1% signal change from baseline.)

resembled that of correct rejections, then visual cortex activity would be congruent with the objective state of the visual world. If, on the other hand, the reverse were true, then visual cortex activity would be congruent with the subject's subjective percept. Drum roll please ... *Figure 19.6.B* indicates that V1 activity was congruent with the percept! The implication for consciousness research, then, is that regions as early as primary sensory cortex may contribute to our conscious awareness of sensory perception.

If you stop to think about this result for a moment, it might be somewhat unsettling for your sense of agency. Is not an implication of the result from false-alarm trials that our visual system may sometimes “show us what we want to see,” rather than what is really out in the world? (I’m reminded of the finding from Soon, Brass, Heinze, and Haynes (2008; *Thought Box 15.1*), that the brain can make a decision long before one becomes aware of it.) Perhaps. But keep in mind that subjects were performing under atypically difficult conditions – low-contrast targets, continuous target presentation, speeded responses; and no feedback. Additionally, isn’t it at least somewhat reassuring to encounter an effect in this, our last chapter together, that reinforces principles that we’ve encountered earlier in the book, in this instance the hypothesis-testing idea articulated by Sillito and colleagues all the way back in *Chapter 3*?

Although the false-alarm trials from Ress and Heeger (2003) might be spookier, for the reasons just discussed, we also need to come up with a satisfactory account for the miss trials from this study. We’ll address this first, then return to the false alarms.

What intrinsic factors influence conscious visual perception?

If one takes a deterministic view of the brain (which we do), there must have been one or more factors that differed in the momentary state of subjects’ brains during miss trials vs. during hit trials from the Ress and Heeger (2003) study. One possibility that was raised at the beginning of the chapter is the spontaneous fluctuation in posterior alpha-band power. The study from Romei et al. (2008; *Research Spotlight 3.2.*) demonstrated that the excitability of visual cortex varies systematically with spontaneous fluctuations of power in the alpha band. A second factor that influences near-threshold visual stimulus detection is the phase of low frequency oscillations during which a stimulus is presented. Mathewson and colleagues (2011) have proposed that oscillations in the alpha band implement a “pulsed inhibition” of neuronal activity, such that the phase of the ongoing alpha-band oscillation at which a stimulus is presented can determine how effectively it is processed. Relatedly, VanRullen, Busch, Drewes, and Dubois (2011) have provided evidence for a ~7 Hz rate of periodic attentional sampling. One interesting exercise would be to revisit this evidence for periodic “up” and “down” states (“open” and “closed”?) of our perceptual awareness after reading the final section of this chapter, on theories of consciousness, and to try to interpret these phenomena in the context of the theories that we’re about to review.

Bottom-up and top-down influences on conscious visual perception

Now we turn to the false-alarm finding from Ress and Heeger (2003), that the V1 “response” to an absent target was almost as robust as its response to real targets, but only on the trials when subjects thought that they had seen a target. One possible explanation is that the conditions of this experiment promoted noisy retino-geniculo-striate signaling, such that on some target-absent trials there was simply more afferent

input to V1 than on others. By this account, target-absent trials with randomly higher input to V1 would have produced greater V1 physiological responses, which, in turn, would trigger false-alarm behavioral responses. This explanation follows the logic that we saw in the microstimulation study of Hanks, Ditterich, and Shadlen (2006) that was described in *Research Spotlight 15.1*. That is, from the perspective of the signal-detection theory of perceptual decision making (*Methodology Box 15.1*), perhaps there was more afferent “momentary evidence” on false-alarm trials than on correct-rejection trials. Although this explanation seems sensible when only considering false-alarm trials, it becomes more problematic when one takes into account the miss trials.

The most straightforward prediction of a “noisy feedforward” account of the Ress and Heeger (2003) data would be that one should see an orderly progression, from little-to-no V1 response on correct-rejection trials, to a small-to-intermediate level of V1 activity on false-alarm trials, to an intermediate-to-large level of V1 activity on miss trials, to a large level

of V1 activity for hits. But this is not what was observed. Instead, the results seem much more consistent with an account whereby intrinsic, possibly top-down, factors suppressed momentary evidence on miss trials and boosted it on false-alarm trials. In *Chapter 3* we considered anatomical and physiological substrates for such top-down influence on V1 (and LGN) processing of an afferent motion signal. Next we consider a study whose results suggest a physiological basis for this top-down account of perceptual awareness.

The role of reentrant signals to V1

For the experiment whose results are shown in *Figure 19.7*, the Dutch team of Supér, Spekreijse, and Lamme (2001) recorded from V1 of monkeys performing a target-detection task with many similarities to that of Ress and Heeger (2003): each trial presented a full-screen “background” of small scratch-like lines all oriented in the same direction, with the to-be-detected “figure” defined by rotating the orientation of every line within a square-shaped region by 90°. The figure

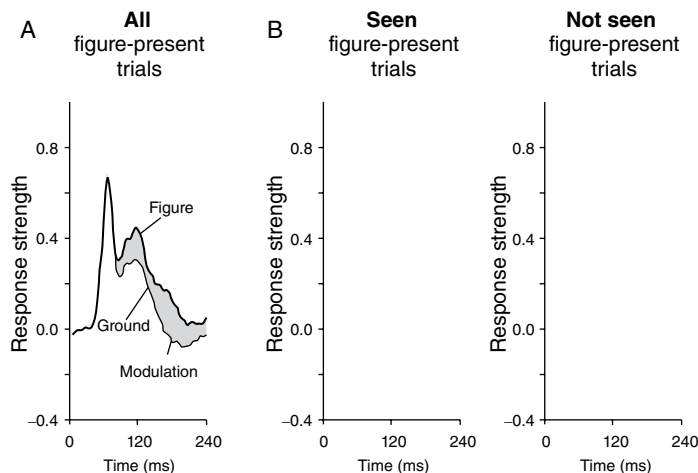


FIGURE 19.7 Normalized spiking activity (“Response strength”) from V1 neurons in a monkey performing a target-detection task. Each trial began with the onset of a full-field textured “background” within which a figure (defined by differently oriented texture) was embedded in one of three locations. These plots show the response of V1 neurons when only the background fell into their receptive fields (“Ground”; light lines) and when the figure fell into their receptive fields (heavy lines). The plot in **A** presents the results from all trials when a figure was present somewhere on the screen (20% of trials, not shown here, were catch trials in which no figure was presented). The plots in **B** show the same data, but broken into hits (“Seen”) and misses (“Not seen”). Source: Supér, Spekreijse, and Lamme, 2001. Reproduced with permission of Nature Publishing Group.

appeared in each of three locations on 26.7% of trials, the remaining 20% being catch trials with no figure. The first thing to notice from *Figure 19.7* is that the neural detection of the figure (i.e., the “modulation”) lagged behind the initial, nonspecific, stimulus-evoked ramp-up of activity by roughly 90 ms. This effect is reminiscent of the effects of object-based attention reported by Chelazzi, Duncan, Miller, and Desimone (1998; *Figure 6.20.C*), in that both show a delayed boost in activity that corresponds to the discrimination between two visual elements in the display. The effect in the Supèr, Spekreijse, and Lamme (2001) data must derive from a different source, however, because its “boost” comes much earlier than the earliest that selective attention effects are seen in V1, which is 200–250 ms after stimulus onset. (Note that the 120–175 ms latencies reported in Chelazzi et al. (1998) were recorded from vAIT.)

The monkeys’ performance was such that there were too few unambiguous false-alarm responses to analyze. What’s shown in *Figure 19.7*, panel B, however, is that activity on miss (i.e., “Not seen”) trials completely lacks any evidence of the figure-related modulation. (Thus, had Supèr, Spekreijse, and Lamme (2001) portrayed their data in a manner analogous to how Ress and Heeger (2003) did theirs, by subtracting baseline (i.e., “Ground”) from target-present (i.e., “Figure”) trials, V1 activity on the miss (i.e., “Not seen”) trials from *Figure 19.7*, panel B, would also have been a flat line with a value of 0.) Supèr, Spekreijse, and Lamme (2001) suggest that the source of figure detection-related modulation of V1 was a reentrant (i.e., feedback) signal coming from a higher station in extrastriate cortex. We will also return to this idea in the final section on theories of consciousness.

Next, we turn to a study whose logic relies on extrinsic, rather than intrinsic, factors to control whether or not subjects are aware of a given stimulus.

Manipulating extrinsic factors to study conscious vs. unconscious vision

The French cognitive neuroscientist Stanislaus Dehaene has made important contributions to at least three subdisciplines within the field: numerical cognition,

reading (see *Further Reading* list from *Chapter 18*), and consciousness. The study that we’ll consider here combines the latter two interests. In it, Dehaene et al. (2001) explored the phenomenon that words presented very briefly and backward masked, such that they are not consciously perceived, can nonetheless produce robust repetition priming effects. (This is reminiscent of the emotional face presentation studies that we considered in *Chapter 17*.) While in the fMRI scanner, subjects viewed a serial stream of jumbled masks, each for a duration of 71 ms, with a mask periodically replaced by one or two 71 ms blank screens. Embedded in these streams were words presented for 29 ms that were either immediately preceded and followed by a blank screen (“unmasked”), or immediately preceded and followed by a mask.

Behavioral data from Dehaene et al. (2001) indicated that masked words went undetected on > 99.3% of trials, whereas unmasked words were detected on > 90% of trials. Despite subjects’ lack of awareness that masked words had been presented, there was evidence in both the behavioral and fMRI data that they were processed to at least the level of the mental lexicon. The behavioral evidence came in the form of repetition priming: when a masked word was followed by the 500 ms (i.e., clearly visible) presentation of a target word about which subjects had to make a “natural/manmade” judgment, responses were significantly faster when the target word was the same as had been the masked word. This was even true when the masked word was presented in an ALL CAPS typeface and the target word in lower case, meaning that it was a relatively high-level representation of the word that had been accessed, not just a low-level visual representation of what the letters physically looked like. The fMRI evidence that masked words were processed to a fairly high level came from the fact that repetition suppression (*Methodology Box 12.2*) was observed in left fusiform gyrus, in the “visual word form area” (labeled “word selective” in *Figure 9.5*), for 500 ms target words that were preceded by masked presentation of the same word, relative to 500 ms target words preceded by masked presentation of a different word.

As illustrated in *Figure 19.8*, the fMRI response evoked by unmasked 29 ms words was markedly

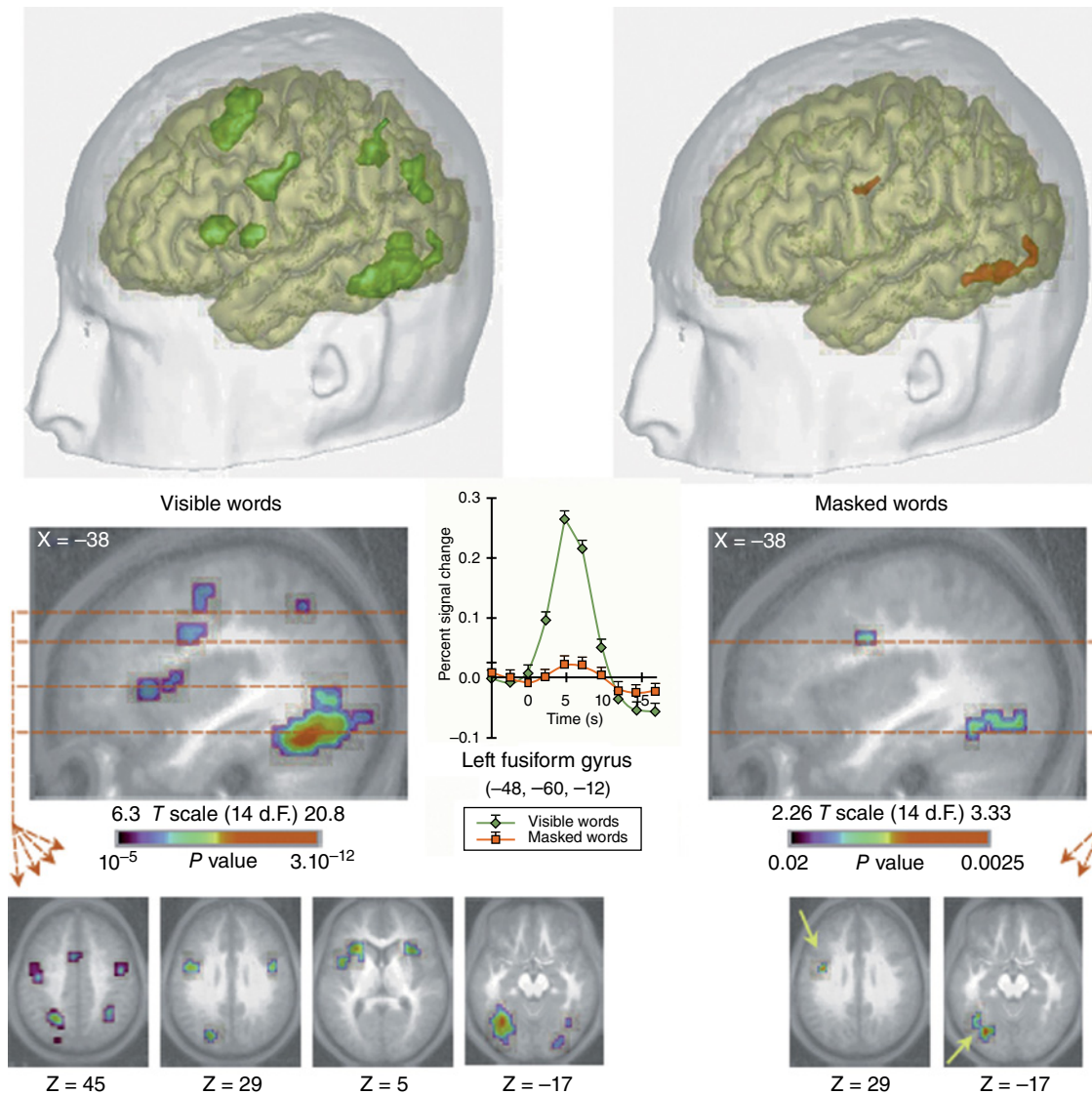


FIGURE 19.8 fMRI activity associated with words that were unmasked, and therefore visible (left-hand column), vs. words that were masked, and therefore not detected (right-hand column). The plot in the center compares the group-averaged response to visible vs. masked words in left fusiform gyrus. Source: Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline, and Rivière. 2001. Reproduced with permission of Nature Publishing Group.

stronger than that for masked 29ms words, both in terms of signal intensity and in terms of the number of regions responding at above-threshold levels. Unmasked words evoked responses in several regions, including fusiform gyrus, parietal cortex, posterior inferior PFC/anterior insula, anterior cingulate, precentral cortex, and supplementary motor area,

bilaterally. Masked words, in contrast, evoked above-threshold responses only in left fusiform gyrus and in left precentral sulcus. Further, when signal intensity in “commonly activated” areas was compared, the response to masked words was only 8.6% as strong as the response to unmasked words in fusiform gyrus, and 5.2% as strong in left precentral sulcus. Finally, for

many of the regions listed here, functional connectivity was significantly greater during the presentation of unmasked than masked words. In summarizing their findings, the authors foreshadowed one of the theories that we'll be considering in the final section of this chapter: "unmasking the words enabled the propagation

of activation and the ignition of a large-scale correlated cerebral assembly" (Dehaene et al., 2001, p. 757).

As a transition to the consideration of theories of consciousness, *Thought Box 19.1* will address a theoretical question, which is the relationship between attention and consciousness.

THOUGHT BOX

19.1 Are attention and consciousness the same thing?

My intuition certainly tells me that they are. I am aware of that to which I pay attention and, conversely, how would it be possible for me to be aware of something to which I'm not attending? The classic example is one considered earlier in the book: the feeling of the pressure of the chair that you're currently sitting in pressing against your body. The sensory information coming from your legs, etc., was no different 10 seconds ago than it is now, yet without attending to it, you weren't aware of it. But there are at least two really smart groups of people who have given this question considerable thought, and who would disagree.

Neuroscientist Victor Lamme, whose study of visual awareness in the monkey is featured in *Figure 19.7*, argues that we can be aware of much more information than we can attend to. Whereas the capacity of the number of items we can simultaneously attend to is limited to four (or less), we can be consciously aware of much more information. One example that he points to is iconic memory. The classic demonstration is from the psychologist George Sperling (1960), in which a 4 x 3 array of letters is briefly flashed (50 ms). If subjects are asked to write down as many letters as they can remember seeing (the so-called "full report" procedure), they typically only recall three or four items. If, however, after the offset of the array, a high-, medium-, or low-pitched tone is played, indicating which row is to be "read out" from memory, subjects will typically remember three items from that row. Because they did not know prior to the tone which row would be cued, it must be the case that, for that brief moment, subjects had at least nine items in mind. Part of Lamme's (2003) argument is that attentional selection is necessary for any kind of overt response, and that it's the attentional selection of items from conscious awareness into the focus of attention that is the capacity-limiting step, during which the representation of nine items shrinks to three. More generally, he argues that attentional selection is the process whereby one (or a few) mental representations that are competing to control behavior get selected for that purpose. Conscious awareness of visual representations arises through other processes (to be considered in the next section) independent of whether or not they will be selected by attention to influence behavior.

Tsuchiya and Koch (2009) also argue that these two constructs can be shown to operate independently. An example of attention without consciousness comes from a variant of the task used by Dehaene et al. (2001). In it, Naccache, Blandin, and Dehaene (2002) demonstrate that the priming for masked words (i.e., for words for which subjects have no awareness) only occurs when they are attending to the display. As an example of consciousness without attention, they point to dual-task studies in which subjects' attention is focused on a central task, yet they are able to identify each time a briefly flashed image in the periphery contains an animal, information that they became aware of without attending to it (Li, VanRullen, Koch, and Perona, 2002).

THEORIES OF CONSCIOUSNESS

It is remarkable that that most of the work in both cognitive science and the neurosciences makes no reference to consciousness (or “awareness”), especially as many would regard consciousness as the major puzzle confronting the neural view of the mind and indeed at the present time it appears deeply mysterious to many people. This attitude is partly a legacy of behaviorism and partly because most workers in these areas cannot see any useful way of approaching the problem. (Crick and Koch, 1990, p. 263)

A great deal has changed in the quarter century since those words were written. Indeed, much of the early energy, the ideas, and the cheerleading driving consciousness research came from this duo, the Nobel Laureate molecular biologist, co-discoverer of the structure of DNA, and the young computational neuroscientist. Today, in contrast to 1990, there is an Association for the Scientific Study of Consciousness, and a handful of journals devoted to the topic.

In this final section, we’ll summarize just a small number of neural theories of consciousness, working backwards through the chapter up to this point, to highlight how theoretical and empirical work on this problem are interwoven. (Note that one realm of theorizing about consciousness that we won’t be covering here, its evolutionary bases, is considered at *Web Link 19.2*.)

Global Workspace Theory

The Global Workspace Theory was first developed by Bernard Baars, based on global workspace computer architecture with a distributed set of knowledge sources, which could cooperatively solve problems that no single constituent could solve alone. Considerable theoretical and empirical work with this model has been carried out by Dehaene and his colleagues. In 2003 Dehaene, Sergent, and Changeux published a neural network model proposing that “the step of conscious perception, referred to as access awareness, is related to the entry of processed visual stimuli into a global brain state that links distant areas including the prefrontal cortex through reciprocal connections, and thus makes perceptual information reportable by multiple means” (p. 8520). This step of access awareness

corresponds to the “ignition of a large-scale correlated cerebral assembly” triggered by the presentation of unmasked words in the Dehaene et al., (2001) study that we considered in the previous section.

A key component of the model is competition among stimulus representations for representation in the global workspace. The global workspace itself is implemented as high-level cortical areas (representing prefrontal, parietal, temporal, and cingulate) that implement “brain-scale propagation of stimulus information,” which results in its accessibility for consciousness. There are several distinctive features to this model. A central tenet that differentiates it from many other models is the all-or-none property of conscious awareness of a representation:

According to the present model, the primary correlate of conscious access is a sudden self-amplifying bifurcation leading to a global brain-scale pattern of activity. This leads to several critical predictions for neuroimaging and pharmacological experiments. First, in experiments with a continuously varying and increasingly visible stimulus, we predict a sudden nonlinear transition toward a state of globally increased brain activity (“ignition”), particularly evident in prefrontal, cingulate, and parietal cortices, accompanied by a synchronous amplification of posterior perceptual activation and by thalamo-cortical brain-scale synchrony in the [gamma] range (20–100 Hz). (Dehaene, Sergent, and Changeux, 2003, p. 8525)

This “all-or-noneness” is illustrated in *Figure 19.9*, in which the simulated attentional blink effectively prevents any representation of T2 in the global workspace. Further, in simulations with the model, when long-distance top-down connections within areas C and D of the global workspace and between C and D and earlier, sensory levels of the model were removed, weakened by 50%, or reduced to only short-range connections, the attentional blink no longer occurred (i.e., T2 gained access to the global workspace). An empirical demonstration that the authors point to is the finding illustrated in *Figure 19.8*: unmasked words, which were consciously perceived, evoked robust activity in a broadly distributed array of cortical areas; masked words, in contrast, of which patients were unaware, evoked relatively weaker activity, and only in a narrow set of regions that is automatically engaged by the visual presentation of words.

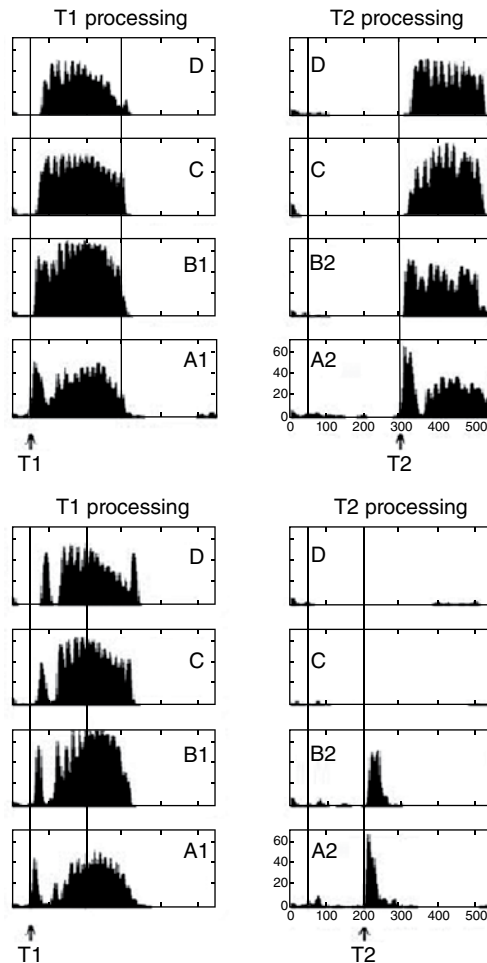


FIGURE 19.9 Simulation of two trials from the attentional blink task in a neural network model of Global Workspace Theory. In both groups, each rectangle corresponds to a level in the model, A1 and B1 being sensory channels that process the T1 stimulus; A2 and B2 being sensory channels that process the T2 stimulus. C and D are higher level areas in which the stimulus must compete for representation. Activity within each level represents firing rate. The trial illustrated in the upper panels simulates a long lag (250ms) between the two stimuli, and as a result, T2 is presented at a time when the sustained activity evoked by T1 has decayed. T2 is therefore able to invade all cortical levels and elicits long-lasting activity comparable to T1. The trial illustrated in the lower panels simulates a “blink” trial that featured a shorter (150ms) T1–T2 lag. Because T1 is still dominating the workspace when T2 arrives, T2 is blocked from entering, and therefore fails to evoke neural activity beyond an initial bottom-up activation in areas A and B. Source: Dehaene, Sergent, and Changeux, 2003. Reproduced with permission of PNAS Group.

Recurrent Processing Theory

Victor Lamme, whose group’s empirical work is illustrated in *Figure 19.7*, has assembled a theoretical account from his analysis of the neurophysiology of visual perception:

When a new image hits the retina, it is processed through successive levels of visual cortex, by means of feedforward connections, working at an astonishing speed. Each level takes only 10ms of processing, so that in about 100–150ms the whole brain “knows” about the new image before our eyes, and potential motor responses are

prepared. From the very first action potentials that are fired, neurons exhibit complex tuning properties such as selectivity for motion, depth, colour or shape, and even respond selectively to faces. Thus, the feedforward sweep enables a rapid extraction of complex and meaningful features from the visual scene, and lays down potential motor responses to act on the incoming information.

Are we conscious of the features extracted by the feedforward sweep? Do we see a face when a face-selective neuron becomes active? It seems not. Many studies, in both humans and monkeys, indicate that no matter what area of the brain is reached by the feedforward sweep, this in itself is not producing (reportable) conscious experience. What seems necessary for conscious experience is that neurons in visual areas engage in so-called recurrent (or re-entrant or resonant) processing ..., where high- and low-level areas interact. This would enable the widespread exchange of information between areas processing different attributes of the visual scene, and thus support perceptual grouping In addition, when recurrent interactions span the entire sensorimotor hierarchy, or involve the frontoparietal areas, potential motor responses could modify the visual responses, which would form the neural equivalent of task set, attention, etc. (Lamme, 2006, pp. 496–497)

Thus, although the Recurrent Processing and Global Workspace models both emphasize a critical role for recurrent processing, they also differ in important ways, as illustrated in *Figure 19.10*. First, by the Recurrent Processing view, activation by the feedforward sweep is not sufficient, in and of itself, to trigger conscious awareness. For example, the initial feedforward sweep of visual information could activate parietal and frontal oculomotor centers so as to trigger an eye movement toward a target without the subject being consciously aware of the stimulus (*Figure 19.10*, Panel A). Second, as illustrated in *Figure 19.10*, B, recurrent processing that is restricted to ventral-stream processing areas would be sufficient to support conscious awareness in the absence of selective attention. Examples of this type of awareness are highlighted in the arguments raised both by Lamme (2003) and by Tsuchiya and Koch (2009) in *Thought Box 19.1*. Lamme has noted that the distinction between consciousness without attention (*Figure 19.10*, B) and consciousness with attention (*Figure 19.10*, C) maps onto the distinction made by philosopher Ned Block (1996) between “phenomenal awareness” and “access awareness.”

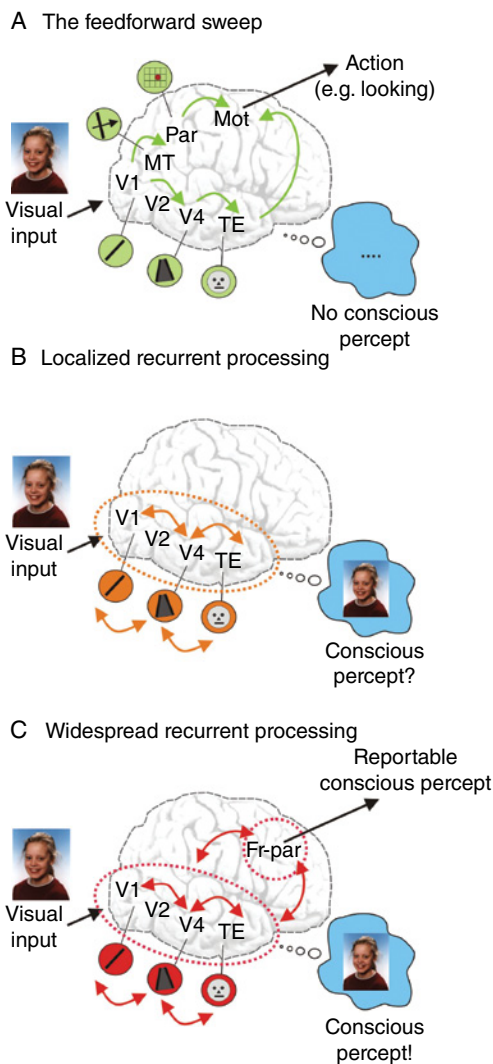


FIGURE 19.10 Three scenarios considered by Lamme’s (2006) Recurrent Processing framework. In **A** the presentation of a stimulus triggers a feedforward sweep along the ventral and dorsal processing streams of the visual system, the icons representing the features that are extracted at each level. **B** corresponds to highly interactive recurrent processing that is nonetheless restricted to the ventral stream. This may correspond to instances of awareness without attention, as described in *Thought Box 19.1*. **C** illustrates a state of widespread recurrent processing, which is equivalent to ignition of the Global Workspace, as specified in Global Workspace Theory. Source: Lamme, 2006. Reproduced with permission of Elsevier.

An important distinction between the Recurrent Processing and Global Workspace models, therefore, is that the former allows for quantitatively graded levels of consciousness, whereas, for the latter, consciousness is an all-or-none proposition. Recurrent Processing Theory also argues, based on neurophysiological evidence, that there can be states in which we are aware of information, but this information is not accessible to overt report. This relates to Lamme's (2006) argument that when behavioral/psychological evidence (e.g., a subject's overt report or performance, or even one's subjective experience) diverges from neurophysiological evidence, the latter should be given equal weight with regard to evidence for conscious awareness of the stimulus in question. Otherwise, he reasons, our construal of consciousness reduces to the cognitive functions necessary for behavior and/or subjective experience – language, motor control, attention, and working memory – and we are no longer studying consciousness per se.

Important challenges for the Recurrent Processing model include specifying what it is about recurrent processing, per se, that gives rise to conscious awareness. Is the activation of certain layers of cortex critical, or the synchronized coactivation of two or more specific layers? Are there different neurotransmitter receptors whose activation is crucial (e.g., AMPA receptors by feedforward glutamate signaling vs. NMDA receptors by feedback)? What are the implications for consciousness of recurrent processing in brain regions whose activity seems to be inaccessible to overt awareness on the part of patients, such as in the right hemisphere of split-brain patients, or in the right hemisphere of patients suffering from unilateral neglect? (Some of these questions, and others, are addressed by Lamme in *Web Link 19.3*.)

Integrated Information Theory

To convey the gist of his Integrated Information Theory (IIT), Giulio Tononi (2008) invites us to consider the following:

You are facing a blank screen that is alternately on and off, and you have been instructed to say “light” when the screen turns on and “dark” when it turns off. A photodiode – a simple light-sensitive device – has also been placed in front of the screen ... The first problem of consciousness reduces to this: when you distinguish between the screen being on or off, you have the

subjective experience of seeing light or dark. The photodiode can also distinguish between the screen being on or off, but presumably it does not have a subjective experience of light and dark. What is the key difference between you and the photodiode?

... the difference has to do with how much information is generated when that distinction is made. Information is classically defined as reduction of uncertainty: the more numerous the alternatives that are ruled out, the greater the reduction of uncertainty, and thus the greater the information ... When the blank screen turns on, ... the photodiode ... reports “light.” ... When you see the blank screen turn on, on the other hand, the situation is quite different ... [because] you are in fact discriminating among a much larger number of alternatives, thereby generating many more bits of information.

This is easy to see. Just imagine that, instead of turning light and dark, the screen were to turn red, then green, then blue, and then display, one after the other, every frame from every movie that was ever produced. The photodiode, inevitably, would go on signaling whether the amount of light for each frame is above or below its threshold: to a photodiode, things can only be one of two ways ... For you, however, a light screen is different not only from a dark screen, but from a multitude of other images, so when you say “light,” it really means this specific way versus countless other ways, such as a red screen, a green screen, a blue screen, this movie frame, that movie frame, and so on for every movie frame (not to mention for a sound, smell, thought, or any combination of the above). Clearly, each frame looks different to you, implying that some mechanism in your brain must be able to tell it apart from all the others. So when you say “light,” whether you think about it or not (and you typically won't), you have just made a discrimination among a very large number of alternatives. (pp. 217–218)

And so, according to this theory, effective connectivity is key, because the number of potential alternative states that the brain can take on determines the amount of information contained in the brain's present state. Tononi calls this measure of integrated information “phi,” and quantitative models for computing the value of phi can be found in Balduzzi and Tononi (2008; 2009). Many of the facts and findings that were reviewed in the first section of this book can be interpreted from the perspective of Integrated Information Theory. Instead of running through these, however, we'll conclude with a set of reanalyses of data from some of these, plus other studies, that demonstrate how something so seemingly abstract

and recondite as a quantitative theory of consciousness can have practical applications.

A practical application of Integrated Information Theory

A problem that we considered earlier in this chapter, and highlighted in *Real-World Applications 19.1* and *Research Spotlight 19.1*, is determining whether patients diagnosed with UWS truly lack all conscious awareness. A multinational team co-led by Brazilian physicist-turned-neuroscientist Adenauer Casali and Belgian neuropsychologist Olivia Gosseries, together with colleagues based in Liège, Milan, and Madison, has begun to address this problem by applying the following rationale:

Phenomenologically, each conscious experience is both differentiated – that is, it has many specific features that distinguish it from a large repertoire of other experiences – and integrated – that is, it cannot be divided into independent components. Neurophysiologically, these fundamental properties of subjective experience rely on the ability of multiple, functionally specialized areas of the thalamocortical system to interact rapidly and effectively to form an integrated whole ... Hence, an emerging idea in theoretical neuroscience is that consciousness requires an optimal balance between functional integration and functional differentiation in thalamocortical networks – otherwise defined as brain complexity ... This kind of complexity should be high when consciousness is present

and low whenever consciousness is lost in sleep, anaesthesia or coma. (Casali, Gosseries, et al., 2013, n.p.)

To assay IIT-inspired complexity in real brains, Casali, Gosseries, et al. (2013) developed a measure inspired by ϕ , called the Perturbational Complexity Index (PCI). From one perspective, it's an approach that's similar in spirit to the “dimensional activation” estimate of complexity in the EEG that was used by Velly et al. 2007 (*Figure 19.4*). However, because the PCI is derived from TMS-evoked signal in the EEG, it is (a) potentially more sensitive (because perturbation of a system with an exogenous force can sometimes reveal properties of a system that aren't evident just by observing it), and (b) not as susceptible to variability in the state of the brain at the time of measurement (because it is computed as a difference score between baseline EEG and the TMS-evoked response in the EEG). (For example, if an individual is experiencing pain or not experiencing pain at the time of EEG measurement, these measurements will be quite different. However, $[TMS/EEG_{\text{pain}} - EEG_{\text{pain}}]$ is likely to be comparable to $[TMS/EEG_{\text{no pain}} - EEG_{\text{no pain}}]$.) “[O]perationally PCI is defined as the normalized Lempel-Ziv complexity of the spatiotemporal pattern of cortical activation triggered by a direct TMS perturbation,” and those who want to dig into the details are directed to Casali, Gosseries, et al. (2013), as well as to *Web Clip 19.2*.

Figure 19.11 illustrates the PCI derived from patients suffering from the neurological syndromes

FIGURE 19.11 The Perturbational Complexity Index applied to anesthesia, sleep, and neurological syndromes of consciousness. Source: Casali, Gosseries, Rosanova, Boly, Sarasso, Casali, Casarotto, Bruno, Laureys, Tononi, and Massimini, 2013. Reproduced with permission of AAAS.

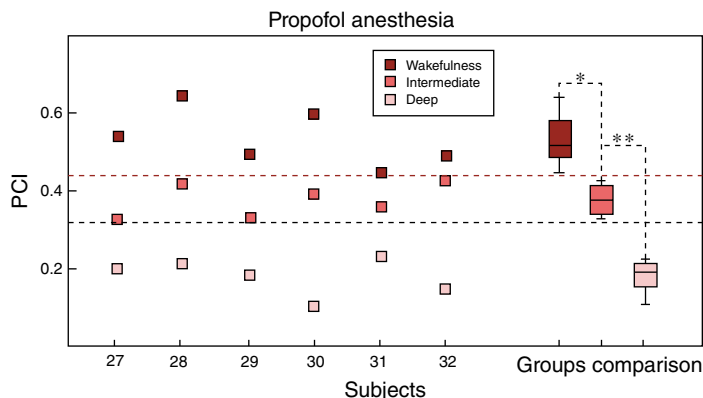


FIGURE 19.11.A PCI is sensitive to graded changes in the level of propofol anesthesia; data from six individual subjects.

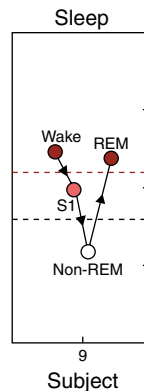


FIGURE 19.11.B PCI is sensitive to changes in sleep state, as illustrated in a single subject transitioning from wakefulness to sleep stage N1 (S1) to nonREM sleep to REM sleep.

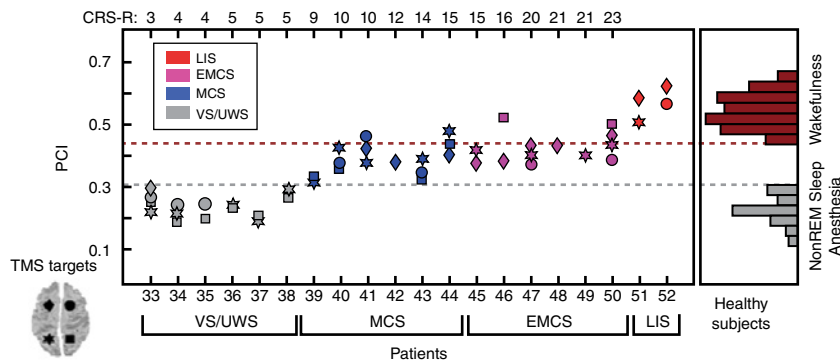


FIGURE 19.11.C PCI, estimated from TMS delivered to each of the four locations indicated on the legend, discriminates vegetative state/unresponsive wakefulness syndrome (VS/UWS) from minimally conscious state (MCS) and emerging-from-MCS (EMCS) from locked-in syndrome (LIS). For comparison, to the right, are histograms showing distribution of PCI scores from healthy individuals in nonREM sleep or under deep general anesthesia, and in wakefulness.

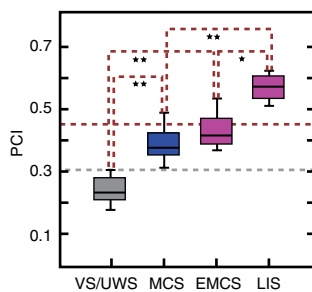


FIGURE 19.11.D Group data from the four neurological groups.

that we reviewed in the first section of this chapter, as well as from individuals in different states of consciousness produced by anesthesia or by sleep. Perhaps most impressively, and most promisingly, it shows the PCI to be sensitive to different levels of consciousness, as associated, for example, with pre-anesthesia wakefulness vs. intermediate levels of anesthesia vs. deep levels. This holds out the promise of a clinically practical procedure, and one far less expensive than fMRI, to objectively, quantitatively, confirm the diagnosis, for example, of persistent vegetative state/UWS.

UPDATING THE CONSCIOUSNESS GRAPH

At the conclusion of this chapter, although we haven't definitively answered the question *What is the Biological Basis of Consciousness?* (not to mention the origin of the universe; or the meaning of life ...), we are, at least, in a position to put some more satisfactory labels on the vertical axis of our two-dimensions-of-consciousness

graph (*Figure 19.2*): the level of awareness seems to depend on connectivity, particularly of lateral and feedback projections, and the complexity of cortical activity that these connections underlie. Of course, there's so much more that we don't yet understand. But amazing discoveries about the neural bases of cognition are being made practically every day. And who knows? Perhaps, gentle reader, it will be you who one day solves the puzzle of the biological bases of consciousness.

END-OF-CHAPTER QUESTIONS

1. What factors – physiological and behavioral – differentiate coma from persistent vegetative state/unresponsive wakefulness syndrome? From locked-in syndrome?
2. Describe the role of the reticular activating system in physiological arousal. How does its activity vary in wakefulness vs. nonREM vs. REM sleep?
3. Are the effects of general anesthesia – physiologically and behaviorally – more akin to coma or to sleep? What physiological change occurs abruptly at the point of anesthesia-induced loss of consciousness?
4. How have scientists addressed the question of whether activity in early sensory cortex more closely reflects events in the outside world or our subjective experience of events in the outside world? And what have they learned?
5. Two sets of studies have emphasized two distinct kinds of physiological response that seem to correspond to whether or not one will be consciously aware of a visually presented stimulus; summarize these.
6. What are the key tenets of Global Workspace Theory? Of Recurrent Processing Theory? Of Integrated Information Theory? Is it possible that more than one of these could be “right”?
7. What are the biological bases of consciousness?

REFERENCES

- Abrams, Daniel A., Srikanth Ryali, Tianwen Chen, Evan Balaban, Daniel J. Levitin, and Vinod Menon. 2013. “Multivariate Activation and Connectivity Patterns Discriminate Speech Intelligibility in Wernicke's, Broca's, and Geschwind's Areas.” *Cerebral Cortex* 23 (7): 1703–1714. doi: 10.1093/cercor/bhs165.
- Balduzzi, David, and Giulio Tononi. 2008. “Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework.” *PLoS Computational Biology* 4 (6): e1000091. doi: 10.1371/journal.pcbi.1000091.
- Balduzzi, David, and Giulio Tononi. 2009. “Qualia: The Geometry of Integrated Information.” *PLoS Computational Biology* 5 (8): e1000462. doi: 10.1371/journal.pcbi.1000462.
- Bauby, Jean-Dominique. 1997. *The Diving Bell and the Butterfly*. Translated by Jeremy Leggatt. New York: Knopf.
- Block, Ned. 1996. “How Can We Find the Neural Correlate of Consciousness?” *Trends in Neurosciences* 19 (11): 456–459. doi: 10.1016/S0166-2236(96)20049-9.
- Casali, Adenauer G., Olivia Gosseries, Mario Rosanova, Mélanie Boly, Simone Sarasso, Karina R. Casali, ... Marcello Massimini. 2013. “A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior.” *Science Translational Medicine* 5 (198): 198ra105. doi: 10.1126/scitranslmed.3006294.
- Chelazzi, Leonardo, John Duncan, Earl K. Miller, and Robert Desimone. 1998. “Responses of Neurons in Inferior Temporal Cortex during Memory-Guided Visual Search.” *Journal of Neurophysiology* 80 (6): 2918–2940. [http://wexler.free.fr/library/files/chelazzi%20\(1998\)%20responses%20of%20neurons%20in%20inferior%20temporal%20cortex%20during%20memory-%20guided%20visual%20search.pdf](http://wexler.free.fr/library/files/chelazzi%20(1998)%20responses%20of%20neurons%20in%20inferior%20temporal%20cortex%20during%20memory-%20guided%20visual%20search.pdf).

- Crick, Francis, and Christof Koch. 1990. "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2: 263–275. <http://authors.library.caltech.edu/40352/1/148.pdf>.
- Dehaene, Stanislas, Lionel Naccache, Laurent Cohen, Denis Le Bihan, Jean-François Mangin, Jean-Baptiste Poline, and Denis Rivière. 2001. "Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming." *Nature Neuroscience* 4 (7): 752–758. doi: 10.1038/89551.
- Dehaene, Stanislas, Claire Sergent, and Jean-Pierre Changeux. 2003. "A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data during Conscious Perception." *Proceedings of the National Academy of Sciences of the United States of America* 100 (14): 8520–8525. doi: 10.1073/pnas.1332574100.
- Ferrarelli, Fabio, Marcello Massimini, Simone Sarasso, Adenauer Casali, Brady A. Riedner, Giuditta Angelini, ... Robert A. Pearce. 2010. "Breakdown in Cortical Effective Connectivity during Midazolam-Induced Loss of Consciousness." *Proceedings of the National Academy of Sciences of the United States of America* 107 (6): 2681–2686. doi: 10.1073/pnas.0913008107.
- Hanks, Timothy D., Jochen Ditterich, and Michael N. Shadlen. 2006. "Microstimulation of Macaque Area LIP Affects Decision-Making in a Motion Discrimination Task." *Nature Neuroscience* 9 (5): 682–689. doi: 10.1038/nn1683.
- Heathcote, Elizabeth. 2011. "Emery Brown: 'Aspects of Anaesthesia Are Consistent with Patients in a Coma.'" *The Observer*, April 9. <http://www.theguardian.com/technology/2011/apr/10/anaesthesia-coma-sleep-emery-brown>.
- Lamme, Victor A. F. 2003. "Why Visual Attention and Awareness Are Different." *Trends in Cognitive Sciences* 7 (1): 12–18. doi: 10.1016/S1364-6613(02)00013-X.
- Lamme, Victor A. F. 2006. "Towards a True Neural Stance on Consciousness." *Trends in Cognitive Sciences* 10 (11): 494–501. doi: 10.1016/j.tics.2006.09.001.
- Laureys, Steven. 2005. "The Neural Correlate of (Un) Awareness: Lessons from the Vegetative State." *Trends in Cognitive Sciences* 9 (12): 556–559. doi: 10.1016/j.tics.2005.10.010.
- Laureys, Steven, Gastone G. Celesia, Francois Cohadon, Jan Lavrijsen, José León-Carrión, Walter G. Sannita, ... European Task Force on Disorders of Consciousness. 2010. "Unresponsive Wakefulness Syndrome: A New Name for the Vegetative State or Apallic Syndrome." *BMC Medicine* 8 (1): 68. doi: 10.1186/1741-7015-8-68.
- Laureys, Steven, Frédéric Pellas, Philippe Van Eeckhout, Sofiane Ghorbel, Caroline Schnakers, Favier Perrin, ... Serge Goldman. 2005. "The Locked-In Syndrome: What Is It Like to Be Conscious but Paralyzed and Voiceless?" *Progress in Brain Research* 150: 495–511. doi: 10.1016/S0079-6123(05)50034-7.
- Laureys, Steven and Giulio Tononi, eds. 2009. *The Neurology of Consciousness*. Amsterdam: Elsevier.
- Li, Fei Fei, Rufin VanRullen, Christof Koch, and Pietro Perona. 2002. "Rapid Natural Scene Categorization in the Near Absence of Attention." *Proceedings of the National Academy of Sciences of the United States of America* 99 (14): 9596–9601. doi: 10.1073/pnas.092277599.
- Massimini, Marcello, Fabio Farrarelli, Reto Huber, Steve K. Esser, Harpreet Singh, and Giulio Tononi. 2005. "Breakdown of Cortical Effective Connectivity during Sleep." *Science* 309 (5744): 2228–2232. doi: 10.1126/science.1117256.
- Mathewson, Kyle E., Alejandro Lleras, Diane M. Beck, Monica Fabiani, Tony Ro, and Gabriele Gratton. 2011. "Pulsed Out of Awareness: EEG Alpha Oscillations Represent a Pulsed-Inhibition of Ongoing Cortical Processing." *Frontiers in Psychology* 2: 99. doi: 10.3389/fpsyg.2011.00099.
- Monti, Martin M., Audrey Vanhaudenhuyse, Martin R. Coleman, Melanie Boly, John D. Pickard, Luaba Tshibanda, ... Steven Laureys. 2010. "Willful Modulation of Brain Activity in Disorders of Consciousness." *New England Journal of Medicine* 362 (7): 579–589. doi: 10.1056/NEJMoa0905370.
- Naccache, Lionel, Elise Blandin, and Stanislas Dehaene. 2002. "Unconscious Masked Priming Depends On Temporal Attention." *Psychological Science* 13 (5): 416–424. doi: 10.1111/1467-9280.00474.
- Owen, Adrian M., Martin R. Coleman, Melanie Boly, Matthew H. Davis, Steven Laureys, and John D. Pickard. 2006. "Detecting Awareness in the Vegetative State." *Science* 212 (5792): 1402. doi: 10.1126/science.1130197.
- Plum, Fred, and Jerome B. Posner. 1983. *The Diagnosis of Stupor and Coma*. Philadelphia: F.A. Davis.
- Ress, David, and David J. Heeger. 2003. "Neural Correlates of Perception in Early Visual Cortex." *Nature Neuroscience* 6 (4): 414–420. doi: 10.1038/nn1024.
- Romei, Vincenzo, Verena Brodbeck, Christoph Michel, Amir Amedi, Alvaro Pascual-Leone, and Gregor Thut. 2008. "Spontaneous Fluctuations in Posterior [alpha]-Band EEG Activity Reflect Variability in Excitability of Human Visual Areas." *Cerebral Cortex* 18 (9): 2010–2018. doi: 10.1093/cercor/bhm229.
- Rosanova, Mario, Olivia Gosseries, Silvia Casarotto, Mélanie Boly, Adenauer G. Casali, Marie-Aurélien Bruno, ... Marcello Massimini. 2012. "Recovery of Cortical Effective Connectivity and Recovery of Consciousness in Vegetative Patients." *Brain* 135 (4): 1105–1115. doi: 10.1093/brain/awr345.

- 1308–1320. doi: 10.1093/brain/awr340 *Brain* 2012; 135: 1308–1320.
- Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. 2008. “Unconscious Determinants of Free Decisions in the Human Brain.” *Nature Neuroscience* 11 (5): 543–545. doi: 10.1038/nn.2112.
- Sperling, George. 1960. “The Information Available in Brief Visual Presentations.” *Psychological Monographs* 74 (11): 1–29. doi: 10.1037/h0093759.
- Supér, Hans, Henk Spekreijse, and Victor A. F. Lamme. 2001. “Two Distinct Modes of Sensory Processing Observed in Monkey Primary Visual Cortex (V1).” *Nature Neuroscience* 4 (3): 304–310. doi: 10.1038/85170.
- Tononi, Giulio. 2008. “Consciousness as Integrated Information: A Provisional Manifesto.” *Biological Bulletin* 215 (3): 216–242. doi: 10.2307/25470707.
- Tsuchiya, Naotsugu and Christof Koch. 2009. “The Relationship Between Consciousness and Attention.” In *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, edited by Steven Laureys and Giulio Tononi, 63–77. Amsterdam: Elsevier.
- VanRullen, Rufin, Niko A. Busch, Jan Drewes, and Julien Dubois. 2011. “Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability.” *Frontiers in Psychology* 2. doi: 10.3389/fpsyg.2011.00060.
- Velly, Lionel J., Marc F. Rey, Nicolas J. Bruder, François A. Gouvitsos, Tatiana Witjas, Jean Marie Regis, ... François M. Gouin. 2007. “Differential Dynamic of Action on Cortical and Subcortical Structures of Anesthetic Agents during Induction of Anesthesia.” *Anesthesiology* 107 (2): 202–212. doi: 10.1097/01.anes.0000270734.99298.b4.
- Vigand, Phillipe, and Stephane Vigand. 2000. *Only the Eyes Say Yes: A Love Story*. English language edition. New York: Arcade.
- Young, G. Brian. 2009. “Coma.” In *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, edited by Steven Laureys and Giulio Tononi, 137–150. Amsterdam: Elsevier.

OTHER SOURCES USED

- Baars, Bernard J. 2005. “Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience?” *Progress in Brain Research* 150: 45–53. doi: 10.1016/S0079-6123(05)50004-9.
- Gosseries, Olivia, Marie-Aurélié Bruno, Audrey Vanhaudenhuyse, Steven Laureys, and Caroline Schnakers. 2009. “Consciousness in the Locked-In Syndrome.” In *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, edited by Steven Laureys and Giulio Tononi, 191–203. Amsterdam: Elsevier.
- Lau, Hakwan. 2011. “Theoretical Motivations for Investigating the Neural Correlates of Consciousness.” *WIREs Cognitive Science* 2 (1): 1–7. doi: 10.1002/wcs.93.
- Laureys, Steven, Maryliese E. Faymonville, André Luxen, Maurice Lamy, George Franck, and Pierre Maquet. 2000. “Restoration of Thalamocortical Connectivity after Recovery from Persistent Vegetative State.” *Lancet* 355 (9217): 1790–1791. doi: 10.1016/S0140-6736(00)02271-6.
- Laureys, Steven, Christian Lemaire, Pierre Maquet, Christophe Phillips, and George Franck. 1999. “Cerebral Metabolism during Vegetative State and after Recovery to Consciousness.” *Journal of Neurology, Neurosurgery, and Psychiatry* 67 (1): 121–122. doi: 10.1136/jnnp.67.1.121.
- Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.
- Tononi, Giulio. 2009. “Sleep and Dreaming.” In *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, edited by Steven Laureys and Giulio Tononi, 89–107. Amsterdam: Elsevier.

FURTHER READING

- Blake, Randolph, and Frank Tong. 2008. “Binocular Rivalry.” *Scholarpedia* 3 (12): 1578. doi: 10.4249/scholarpedia.1578.
An authoritative and succinct summary of the history of the study of this intriguing phenomenon.
- Brown, Emery N., Ralph Lydic, and Nicholas D. Schiff. 2010. “General Anesthesia, Sleep, and Coma.” *New England Journal of Medicine* 363 (27): 2638–2650. doi: 10.1056/NEJMra0808281.
Authoritative review by prominent clinician-scientists focusing on the neural mechanisms of unconsciousness induced by selected intravenous anesthetic drugs.

- Crick, Francis. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribner.
Francis Crick's influential book is among the publications that made consciousness a respectable topic for study by mainstream cognitive neuroscientists.
- Crick, Francis. 1996. "Visual Perception: Rivalry and Consciousness." *Nature* 379 (6565): 485–486. doi: 10.1038/379485a0.
A succinct statement of why and how the study of binocular rivalry has important implications for understanding the neural correlates of consciousness.
- Engel, Andreas K., Pascal Fries, Peter König, Michael Brecht, and Wolf Singer. 1999. "Temporal Binding, Binocular Rivalry, and Consciousness." *Consciousness and Cognition* 8 (2): 128–151. doi: 10.1006/ccog.1999.0389.
Summary of the influential view that oscillatory synchrony among anatomically distal brain areas is of central importance to conscious experience; also reviews binocular rivalry, a phenomenon that is often used in the study of visual awareness.
- Laureys, Steven, and Giulio Tononi, eds. 2009. *The Neurology of Consciousness*. Amsterdam: Elsevier.
A volume containing chapters by a proverbial "veritable who's-who" of scientific and clinical experts on disorders of consciousness.
- Leopold, David A., and Nikos K. Logothetis. 1996. "Activity Changes in Early Visual Cortex Reflect Monkeys' Percepts during Binocular Rivalry." *Nature* 379 (6565): 549–553. doi: 10.1038/379549a0.
An empirical paper that Crick (1996) declared "among the opening salvos of a concerted attack on the baffling problem of consciousness."
- Nir, Yuval, and Giulio Tononi. 2010. "Dreaming and the Brain: From Phenomenology to Neurophysiology." *Trends in Cognitive Sciences* 14 (2): 88–100. doi: 10.1016/j.tics.2009.12.001.
Overview of this fascinating arena from which to study consciousness.
- Seager, William. 2007. "A Brief History of the Philosophical Problem of Consciousness." In *The Cambridge Handbook of Consciousness*, edited by Philip David Zelazo, Morris Moscovitch, and Evan Thompson, 9–33. New York: Cambridge University Press.
This chapter offers a nice overview for the nonphilosopher.
- Tononi, Giulio. 2012. *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon Books.
A literary, fantastical dream in which Galileo, alternately guided by Francis Crick, Alan Turing, and Charles Darwin, encounters a variety of neurological patients whose syndromes define the parameters of consciousness.

GLOSSARY

acoustic energy The movement of molecules of the gases that make up the air that's all around us; or, when underwater, collisions between and movement of molecules of water.

action A isolated movement, or a coordinated series of movements intended to achieve some behavioral goal.

affective neuroscience The study of the neural bases of emotions. Because disorders of emotional processing are often characteristic of, if not at the root of, many types of psychopathology (e.g., depression, addiction, generalized anxiety disorder), there is considerable overlap between affective neuroscience and clinical psychology. (*Affective* is the adjectival form of the noun *affect*, a synonym for *emotion*.)

allocentric localization The representation of space in an environment-centered coordinate system (e.g., *The cup is on the table*).

anatomical specificity This refers to the degree to which a particular effect on behavior is attributable to only a particular anatomical region. See *Methodology Box 1.1 on Neuropsychology, neurophysiology, and the importance of anatomical specificity*.

anions A general term for negatively charged ions.

anomia An inability to produce the appropriate verbal label for objects (i.e., to “name” things).

anoxia Deprivation of oxygen to brain tissue.

anterior Used as a term of direction, it refers to the front (a synonym, introduced in *Chapter 2*, is “rostral”).

antisaccade task A task requiring a speeded saccade in the opposite direction of a temporally unpredictable target; it requires cognitive control to overcome the prepotent tendency to saccade to unexpected appearance of stimuli in the environment.

aperture problem The small size of receptive fields at early stages of the visual system can result in ambiguities when local information present within the aperture is inconsistent with global information in the scene.

aphasia An inability to use language following brain damage.

attention The process of prioritizing the processing of a subset of the information that is available at any given moment in time.

attention deficit hyperactivity disorder A neurological condition associated with difficulty sustaining attention, difficulty controlling behavior, and, in some cases, hyperactivity.

auditory nerve The bundle of axons that carry auditory information from the cochlear nucleus to the brainstem auditory nuclei.

autism spectrum disorder A developmental disorder characterized by deficits in social communication and social interaction, and often accompanied by restricted, repetitive patterns of behavior, interests, or activities.

avoidance learning Context-dependent skeletomotor behaviors that, when engaged, enable an organism to avoid the adverse outcome predicted by a conditioned stimulus (CS).

axons The branches through which a neuron sends signals to other neurons. Most axons have many “collateral” arms splitting off from the main branch, thereby enabling a single neuron to influence hundreds, thousands, and in some cases tens of thousands of other neurons.

axon hillock Region of the cell body from which the axon emerges. The membrane of the hillock is dense

- with voltage-gated Na^+ channels, such that it can powerfully initiate action potentials.
- axonal transport** Intracellular mechanisms for transporting proteins from the cell body to synapses.
- barrel cortex** The region of rodent somatosensory cortex containing groupings of neurons in layer IV – “barrels” – that, together, represent somatosensory signals coming from a single whisker.
- basilar membrane** The membrane within the cochlea whose vibrations effect the spectral transformation of sound, which is subsequently encoded in the cochlear nucleus’s place code. The base of the basilar membrane is the portion nearest the oval window, and is maximally displaced by high frequencies. The apex is the “far end” of the basilar membrane, and is maximally displaced by low frequencies.
- binding problem** The fact that anatomically discrete parts of the brain encode what an object is vs. where it is, yet our perception is of an object occupying (i.e., “bound to”) a location.
- binocular rivalry** The phenomenon that when a different image is presented to each eye, one sees only one of the two at any given point of time, and the perceived image alternates, stochastically, between the two.
- blood–brain barrier** The layer of endothelial cells lining the inside walls of blood vessels that prevents the diffusion from capillaries of many substances carried in the blood stream.
- blood oxygen level-dependent (BOLD) signal** The signal most commonly detected by fMRI, measured via T2^* relaxation.
- brain** The organ of the body responsible for the control of all other systems of the body, and thus for the control of behavior. In technical, anatomical terms, the brain is comprised of the brainstem, cerebellum, and cerebrum (see *Chapter 2*).
- brainstem auditory nuclei** The many subcortical structures that process the auditory signal prior to the inferior colliculus, medial geniculate nucleus, and primary auditory cortex.
- brainstem oculomotor nuclei** Neurons located at the rostrocaudal level of the superior colliculus whose axons innervate the muscles that control the position and movements of the eyes (the “extraocular muscles”).
- cations** A general term for positively charged ions.
- central nervous system** Comprises the brain plus the spinal cord and the motor neurons that leave the cord to innervate the muscles of the body.
- cerebellum** The highly folded structure that lies “below” the “back” of the cerebral cortex. (See *Figure 1.2*; correct terms of direction are introduced in *Chapter 2*.)
- cerebral hemisphere** The *cerebrum* is the largest part of the brain (in volume), sitting atop the brainstem. It is divided into two roughly symmetrical halves, each referred to as a hemisphere.
- cerebrospinal fluid (CSF)** The fluid that circulates through the ventricles (thereby inflating the brain) and also along the exterior surface of the CNS (thereby providing a medium in which the brain floats).
- cerebrum** Also referred to as the forebrain, the cerebrum is the anterior-most part of the brain, and largest in volume. Deriving from the embryonic telencephalon, it comprises the thalamus plus all the tissue that is rostral to it. Major elements include the five lobes of cerebral cortex, basal ganglia, hippocampus, thalamus, and all the white matter connecting these structures.
- cognition** Thinking. This word derives from Greek and Latin words meaning “to know” or “to learn” and, in some translations, “to recognize.”
- cognitive neuroscience** The studies of the neural bases (e.g., anatomical, physiological, pharmacological, genetic) of behavior that have traditionally fallen under the purview of cognitive psychology.
- cognitive psychology** Cognitive psychology is the study of the information processing that intervenes between sensory inputs and motoric outputs. Domains of mental activity and behavior that it covers include those covered in this book, beginning with *Chapter 3*.
- conjugate eye movements** Movements of the two eyes in parallel (i.e., with the same vector).
- consolidation** In the context of long-term memory, consolidation is the hypothesized process whereby, over time, memories become more resistant to disruption.
- construct validity** the extent to which there is a principled basis for asserting the existence of a psychological, or physiological, construct.
- convolution (*n.*, to convolve, *v.*)** The transformation of a function (e.g., a time-varying signal) that results from passing it through a filter.
- coordinate transformations** The conversation of spatial information from one reference frame to another, such as when re-representing a location initially encoded in eye-centered coordinates into head-centered coordinates.
- corpus callosum** (Latin for “tough body,” presumably because, relative to other structures, it is difficult to cut through or pull apart.) The corpus callosum is made up of millions of axons that project between neurons in the two hemispheres.
- cortex** Refers to the layer of tissue containing cell bodies, and thus is often synonymous with “gray matter.” Cortex

- is the Latin name for “rind,” and thus the name references the outer covering of many types of fruit, such as melons or citrus fruit. The cerebrum and cerebellum each have a cortex.
- cortical column** A group of neurons that are stacked, as though in an imaginary tube that is oriented orthogonal to the cortical surface. Neurons within a cortical column are more strongly interconnected, and, therefore, share more functional properties, than are physically adjacent neurons that are not within the same column.
- cortical magnification** The phenomenon that higher-resolution regions of a sensory-transduction surface, or more finely controlled muscles, are represented by larger areas of cortex.
- cotransport** Process whereby more than one molecule is simultaneously carried across a cell membrane, as with reuptake.
- covert attention** Attention deployed toward an object or location without looking directly at it.
- craniotomy** Surgical procedure of removing a portion of the cranium (skull) in order to expose a portion of the underlying brain tissue.
- decision variable** A hypothesized representation of all the accruing information related to a decision.
- declarative memory** Long-term memory that entails conscious retrieval of information from one’s past. Synonymous with *explicit memory*.
- degrees of visual angle** When foveating an object in the environment, the surface area of retina that that object occupies can be expressed as an angle anchored in the fovea, and whose imaginary sides project to the left and right edges of the object. The larger the object, the wider the degrees of visual angle that it occupies.
- delay discounting** The tendency to assign less value to a reward if its delivery will occur at some point in the future, rather than immediately.
- dementia** A progressive impairment of intellect and behavior that, over time, dramatically alters and restricts normal activities of daily living. Although typically implicating memory, dementia differs from amnesia in that it also affects at least one other domain of cognition (e.g., language, motor control, visual perception, ...).
- dendrites** The branches of a neuron that receive signals from other neurons (*Figure 2.7*).
- depolarize** To make the membrane potential of a cell more positive. Typically, because the resting state of neurons is negative, this is equivalent to making the membrane potential less negative.
- diencephalon**, also commonly referred to as the **thalamus** Large collection of nuclei located near the center of the cerebrum, straddling the midline and immediately rostral (see *Figure 1.2*) to the midbrain. Many thalamic nuclei serve as relay points between peripheral sensory information and the cerebral cortex.
- diffusion tensor imaging** A type of MRI that measures the diffusion of water to infer information about white-matter tracts.
- dipole** An object with a positive charge at one end and a negative charge at the other.
- distributed system** A system in which functions arise from the contribution of many processing elements performing simple computations (i.e., carrying out simple functions) that often bear little surface resemblance to the function in question. These elements may or may not be anatomically distant from each other.
- double dissociation of function** A pattern of results in which damage to area A affects performance on task X, but not on task Y; whereas damage to area B affects performance on task Y, but not on task X.
- dynamics** The forces and torques that produce movement or motion.
- dysphasia** An impairment in the ability to communicate following brain damage (compare with *aphasia*).
- efference copy** A copy of the motor command that is carried by a recurrent circuit.
- egocentric localization** The representation of space in a body-centered coordinate system (e.g., *The cup is to my right*).
- electrocorticography (ECoG)** Electrical recordings made from the surface of the cortex.
- emergent property** A property that is characteristic of a system, but that is not reducible to the properties of the system’s constituent elements.
- encoding** In the context of memory, encoding refers to the processing of newly encountered information such that it can be stored and later retrieved.
- engram** The hypothetical physical representation of a memory in the brain.
- ensembles** From the French for “together,” this term is often used to refer to a group of neurons that are “working together” as a functional unit. The analogy is to a musical ensemble, such as a string quartet.
- epilepsy** A chronic condition of repeating seizures.
- epileptogenic** Brain tissue with a low threshold for initiating seizure activity.
- episodic memory** Declarative memory that includes a representation of the context (e.g., place, situation, who

- was present) at the time that encoding of that information occurred.
- equipotentiality** The idea that any given piece of cortical tissue has the potential to support any brain function.
- etiology** The cause of a disease.
- excitatory neurotransmitter** A neurotransmitter whose effect on the postsynaptic membrane is to depolarize it.
- explicit memory** Long-term memory that entails conscious retrieval of information from one's past. Synonymous with *declarative memory*.
- express saccades** Saccades with an onset latency that is roughly one half that of standard, volitionally triggered saccades; they are generated in response to the abrupt onset of an unexpected stimulus.
- extinction** A phenomenon often accompanying unilateral neglect, whereby a patient can be aware of a single object presented on her left, but will lose awareness of it if a second item is then presented on her right.
- extracellular recording** Electrical recordings made from an electrode located in extracellular space. Such recordings can be sensitive to action potentials and/or LFPs.
- extrapersonal space** Those portions of perceived space that are further than we can reach.
- extrastriate cortex** Cortex that is part of the visual system, but not a part of striate cortex (i.e., it is beyond visual cortex).
- familiarity** The subjective experience that accompanies correctly recognizing an item on a recognition memory test (or, in real life, maybe a face on the bus) without a memory for the context in which that item was previously experienced. Hypothesized in some models to correspond to a neural signal and/or process that is qualitatively different from recollection.
- feature detector** An element in a sensory processing array whose function is to signal each time it detects the presence of the feature for which it is tuned.
- fMRI adaptation** The experimental procedure of habituating the response of a region to repeated exposure to a stimulus (or class of stimuli), then varying the stimuli by a dimension of interest and measuring whether there is a "release from adaptation," which would indicate that the region in question is sensitive to change along the dimension being varied.
- foraging decision** The more-or-less-continual decision of whether to persist in one's current mode of behavior or whether to change to a different mode of behavior. Compare with *perceptual decision* and *value-based decision*.
- formants** Bands of sustained, elevated oscillatory power in a spectrogram of speech.
- fornix** The bundle of fibers leaving the hippocampus, initially in the caudal direction, and arcing dorsally and rostrally to synapse on targets in the hypothalamus and basal forebrain. It is generally understood that these fibers are important for maintaining the physiological state of the hippocampus (e.g., maintenance and regulation of hippocampal theta rhythm) rather than carrying the "informational output" of the hippocampus, which, instead, is carried by projections to entorhinal cortex.
- forward model** A model that predicts the consequences of an action.
- Fourier transform** The translation of a time-varying wave form into constituent sine and cosine waves.
- fovea** Region of the retina most densely packed with photoreceptors, and thus supporting highest resolution vision.
- frontal lobe** The rostral-most of the four lobes of the cerebral cortex.
- functional connectivity** A state in which the activity of two or more regions is correlated, and they are therefore presumed to be "working together." (Note that the existence of a "structural connection," need not imply that the connected areas are functionally connected in a particular context.)
- functional magnetic resonance imaging (fMRI)** The application of MRI that is sensitive to physiological signals (as opposed to anatomical structure).
- fundamental frequency** The lowest frequency formant of an utterance.
- fundus** The "bottom" of a sulcus, where its two banks meet.
- GABA (adj. GABAergic)** The abbreviated name of γ -aminobutyric acid, the most common inhibitory neurotransmitter in the CNS.
- gamma frequency band** Neural oscillations in the range of 30–80 Hz.
- ganglion (pl. -glia)** Ganglion is a synonym for *nucleus*. The *basal ganglia* are a group of nuclei located near the base of the cerebrum, and anterior and lateral to the thalamus (see *Figure 2.4*). Their function is considered first in *Chapter 7*.
- gating** The process of allowing certain signals to pass through a circuit, and preventing others.
- generalized anxiety disorder (GAD)** A psychiatric condition characterized by incessant, excessive worry.
- gestalt psychology** A school of thought, originating with the study of perception, that emphasized recognition of whole objects as being more important than feature-level processes.
- glial cells** Glial cells (or "glia") is the designation given to cells in the brain that are not themselves neurons, but can be construed as playing supporting roles that are necessary for the normal functioning of neurons.

Two types of glia mentioned in *Chapter 2* are oligodendrocytes (responsible for myelin in the CNS) and astrocytes.

glutamate (adj. glutamatergic) Also referred to as “glutamic acid,” glutamate is the excitatory neurotransmitter released by cortical pyramidal cells, and thus the primary chemical currency for information transmission in the brain.

grand mal epileptic seizure A seizure producing a loss of consciousness and violent muscle contractions (sometimes also referred to as tonic-clonic).

gray matter Informal, but widely used, term for the cortex. Derives from fact that the cortex lacks myelin, and thus appears darker to the naked eye than does the myelin-rich white matter.

gyri (sing. gyrus) From the Latin for “circles,” convex ridges on the surface of the cerebral cortex. Each gyrus is bounded by sulci on each side.

hidden unit layer A layer in a parallel-distributed processing (PDP) model that neither receives direct external input from, nor sends direct output to, the environment.

high-level cognition An informal, but widely used, classification for most mental functions other than sensory perception, motor control, and some aspects of memory and attention. Examples include language, problem solving, reasoning, and meta-cognition (thinking about thinking).

hippocampus A subcortical structure of the cerebrum that, in the primate, is surrounded by the cortex of the temporal lobe, and lies parallel to this lobe’s long axis. The term is Latin for seahorse, of which its shape is suggestive.

homunculus A “little human” in the brain. In the context of somatosensory and motor neuroscience, the homunculus refers to the somatotopic representation of the post- and precentral gyri, respectively; in the context of psychology, the homunculus refers to a hypothetical little person inside the brain who is responsible for one’s cognitive functions.

horizontal meridian The imaginary horizontal line that divides the visual field into upper and lower quadrants.

hypercolumn A hypothesized functional unit of cortical columns in V1, the totality of which represents every possible feature that could be presented to a specific V1 receptive field.

hyperpolarize To make the membrane potential of a cell more negative. (I.e., the opposite of *depolarize*.)

implicit memory The influence of past experience on present behavior that is accompanied by awareness of

neither this influence, nor the past experience.

Synonymous with *nondeclarative memory*.

individual differences The study of factors underlying why individuals in a sample or population differ from each other. In studies of brain–behavior relationships, for example, a between-subject correlation between levels of some neurobiological dependent measure (e.g., neuroimaging signal in Region x) and variation in a behavioral measure (e.g., RT on Task y) is taken as compelling evidence for a role for x in performance of y.

inhibitory interneurons Along with pyramidal cells, the second major class of neurons in the cerebral cortex. Although there are many types of inhibitory interneuron, this book will primarily refer to the generic interneuron: a GABAergic cell with only local projections to nearby pyramidal cells.

inhibitory neurotransmission Of a neurotransmitter whose effect on the postsynaptic membrane potential is to **hyperpolarize** it, i.e., to make it more negative.

inhibitory surround A mechanism for increasing the signal-to-noise ratio of a neural representation by inhibiting activity corresponding to adjacent, but irrelevant, representations.

inner hair cells Cells that are attached to the basilar membrane, the vibration of which causes them to release neurotransmitter onto a neuron of the cochlear nucleus, thereby effecting the transduction of acoustic energy into a neural signal.

innervated Having axons project into, or containing neurons that send axons from.

intention When used in contradistinction to attention, intention refers to the processes required to prepare for (or, sometimes, just to contemplate) performing an action.

interaural intensity difference (IID) The phenomenon that the acoustic energy from a single sound source, if not positioned in direct alignment with the midline of the head, will be of greater intensity for one ear than the other. This provides a useful cue for sound localization.

interaural timing difference (ITD) The phenomenon that the time of arrival of the acoustic energy emanating from a single sound source, if not positioned in direct alignment with the midline of the head, will be quicker for one ear than the other. This provides a useful cue for sound localization.

intertemporal choice A situation in which an individual must choose between immediate vs. delayed reward, and often producing the phenomenon of delay discounting.

- intracellular recording** Electrical recordings made with a hollow-tube electrode that punctures the membrane of a cell (and forms a tight seal between membrane and electrode) so that fluctuations of voltage within a that individual neuron can be measured.
- invariant** The property of not responding to changes along a particular dimension.
- inverse problem** The fact that there is an infinite number of possible source solutions for any pattern of voltages distributed across the scalp.
- isoluminant** A display in which all objects, plus the background, are presented with the same luminance, and so only color information can be used to segment the scene.
- kinematics** The dynamics of movement or motion without consideration of the forces that impel it.
- laminar** “Lamina” is another word for layer; although derived from Latin, the plural form is “laminae” or “laminae.” It is often used as an adjective when referring to layer-specific (i.e., “laminar”) organization within cortex.
- lateral inhibition** When that activity of a neuron in area X inhibits the activity of other neurons in area X; i.e., the source of the inhibition is neither feedback nor feedforward, but “feed-sideways”/ “feed-laterally.”
- lesion** Used as a noun, it refers to either damage or an abnormality. It is a general term, in that it does not specify either the nature or the cause. Thus, a region of dead or otherwise compromised tissue can be referred to as a lesion regardless of whether it was caused by viral encephalitis (swelling), by anoxia (deprivation of oxygen, such as could happen as a result of a stroke), or by the experimental injection of ibotenic acid, a chemical that kills the soma of neurons, but not fibers of passage (see *Chapter 2* for definitions). (Note that each of these causes would produce a different pattern of damage.) For use as a verb, see definition of *neuropsychology* and *Methodology Box 1.1*.
- ligand** A substance, typically a molecule, that binds to a specific site on a protein, thereby triggering some process. In the context of neurotransmission, the ligand is the neurotransmitter that binds with a binding site on the postsynaptic receptor.
- local field potential (LFP)** The oscillating electrical potential recorded from extracellular space that reflects the moment-to-moment aggregation of fluctuating voltages from all dendritic trees in the vicinity of the electrode.
- localization of function** The idea that a behavioral or cognitive function can arise from a discrete (i.e., “localized”) anatomical region of the brain.
- machine learning** A branch of computer science that studies how computer algorithms can learn about structure in a set of data.
- major depressive disorder (MDD)** A mood state characterized by sadness, poor concentration, and loss of interest in daily activities, that can also be accompanied by fatigue, insomnia, and loss of appetite.
- mass action** The idea that behavioral and cognitive functions cannot be understood as arising from discrete anatomical regions of the brain.
- mechanoreceptors** Receptors in the skin that respond to mechanical forces: pressure, texture, flutter, vibration, and stretch.
- mental lexicon** The subset of semantic memory corresponding to all the words that one knows the meaning of.
- motor effector** The part of the body that actually carries out an action – most often used to refer to the limbs and the eyes.
- motor efferent** An efferent fiber is a fiber projecting away from a region (an afferent is a fiber projecting into a region). The term “motor efferents” refers to fibers leaving the cortex and synapsing in the spinal cord.
- myelin** The cholesterol-laden sheath that insulates axons. Myelin is formed by branches from a type of glial cell, called oligodendrocytes, that wraps the axon in several layers.
- Na⁺/K⁺ pump** Also known as “ATPase,” because it’s an enzyme that requires ATP for its operation. The Na⁺/K⁺ pump is a membrane-spanning complex that actively restores the electrochemical gradients of Na⁺ and K⁺ that are decreased by the mechanisms of the action potential and of neurotransmitter reuptake.
- natural language** Any human language commonly used for verbal or written communication that has arisen in an unpremeditated fashion. Computer-programming languages and languages used in formal logic are not natural languages.
- neural code** The way that the nervous system represents information via action potentials and neurochemicals.
- neurodegenerative disease** A disease characterized by progressive loss of brain tissue.
- neuroeconomics** The study of the neural bases of value-based decision making.
- neurons** The cells in the brain that generate electrical and chemical signals that control all other systems of the body. The typical neuron has four distinct regions: dendrites, cell body, axon, and synaptic terminals (see *Chapter 2*).

- neurophysiology** The study of the functioning of the brain. Two broad classes of neurophysiology are *stimulation* and *measurement*; examples of the latter are introduced in *Chapter 2* and *Chapter 3*.
- neuropsychology** The careful analysis of the behavioral consequences resulting from damage to a particular brain structure. Neuropsychology can either be conducted by experimentally producing a surgical lesion in an animal, or by observing the effects on behavior of brain damage incurred by humans (see *Methodology Box 1.1*).
- nigrostriatal tract** The bundle of axons, originating from the substantia nigra pars compacta, responsible for the release of dopamine onto the neostriatum.
- nondeclarative memory** The influence of past experience on present behavior that is accompanied by awareness of neither this influence, nor the past experience. Synonymous with *implicit memory*.
- nucleus (pl. -clei)** In the context of the cell, the nucleus is the central structure containing DNA. In the context of gross anatomy, a nucleus is a subcortical group of neurons, often visible to the naked eye, that typically all perform the same function.
- ontogeny** The development of an organism; in the case of vertebrates, from fertilization through birth and the entirety of the lifespan.
- outer hair cells** Cells anchored at one end to the basilar membrane, and at the other to the cochlea; their coordinated activity can dampen the vibrations of the basilar membrane in precise areas.
- overt attention** Attention deployed toward an object or location by looking directly at it.
- paraphasic speech** Speech including unintended/inappropriate syllables, words, or phrases.
- perception** The identification of important features of the item or information being sensed.
- perceptual decision** A categorical decision about a perceived stimulus (e.g., *Is it moving up or down?*, *Is that tone louder or softer?*). In effect, *What is the objective state of this specific part of my environment?* Compare with *value-based decision* and *foraging decision*.
- peripersonal space** The space surrounding one's body that is within one's grasp.
- peripheral nervous system** The peripheral nervous system (sometimes abbreviated PNS) is made up of all neurons in the body whose cell bodies are located outside of the central nervous system. These include somatosensory (i.e., "touch") neurons whose cell bodies are in the skin and that send signals via axons that project into the spinal cord, and neurons that control autonomic functions of many organ systems of the body.
- perseveration** The persistent selection of a previously reinforced response (or type of response) after having received feedback that it is no longer appropriate.
- phase synchronization** The phase of a sine wave refers to the location within the cycle (e.g., the peak, or the trough). To synchronize the phase of two oscillators is to adjust them such that, whenever one is at a particular phase in its cycle (say, the peak) the other will always be at the same phase in its (this could also be the peak, or it could be any other phase, so long as its predictable).
- phoneme** The smallest unit of sound in a language that, when changed, can change the meaning of an utterance.
- phototransduction** The conversion of energy from the electromagnetic spectrum ("visible light") into neural signals.
- phrenology** The practice, initiated by Franz Josef Gall (1758–1828), of divining mental traits from patterns of convexities and concavities on the skull.
- phylogeny** The evolutionary history of organismal lineage.
- pinwheel** The actual physical layout, discovered with optical imaging, of the cortical columns that accomplish the function that had been proposed for the hypercolumn.
- place code** The scheme whereby the representation of frequency (pitch) in the cochlear nucleus is laid out in a linear fashion, from high to low.
- place of articulation** The physical location in the mouth where the tongue applies the blockage that produces stop consonants.
- plasticity** The property of nervous tissue that its structure and function can modify in response to a change in environmental inputs (or, indeed, to systematic changes of mental activity).
- posterior** Used as a term of direction, posterior refers to the back (a synonym, introduced in *Chapter 2*, is "caudal").
- post-traumatic stress disorder (PTSD)** A pathological condition resulting from traumatic experience. Symptoms can include re-experiencing the traumatic event (including "flashbacks" and nightmares); avoidance symptoms; and hyperarousal (includes being easily startled, and feeling tense).
- power** The magnitude of an oscillatory component in a signal.
- power spectrum** A representation of the magnitude of each frequency resulting from a Fourier (or *spectral*) transform.
- predictive coding** The idea that a core function of the brain is to construct, and constantly update, a model of what's likely to happen in the future.

- premotor theory of attention** The idea that attention may be a consequence of motor planning, covert attention being a motor plan that never gets realized as an actual movement.
- primary visual cortex** The part of the cerebral cortex that is responsible for the sense of sight.
- proprioception** The sense of where different parts of the body are relative to one another.
- pyramidal cells** Glutamate-releasing neurons with pyramid-shaped cell bodies that are the most common excitatory neuron in the neocortex and hippocampus.
- random walk** A mathematical formalization of a process that proceeds via stochastic steps.
- rate code** A neural code whereby a information about a change in some metric property is conveyed by a change in the rate at which a neuron fires.
- receiver operating characteristic (ROC)** A graph plotting hits vs. false alarms as a function of criterion (often operationalized as “confidence”). Deriving from signal detection theory, this measure is intended to capture the strength of a percept (or, in memory research, of a subject’s memory) independent of decision criteria.
- receptive field** The portion of the sensory-transduction surface that is represented by a neuron (see *Thought Box 3.2*).
- recollection** In the context of memory retrieval, the act of remembering the context in which a fact was learned or an event occurred. Hypothesized in some models to correspond to a neural signal and/or process that is qualitatively different from familiarity.
- recurrent circuit** A circuit carrying feedback signals.
- region of interest (ROI)** In neuroimaging, a region from which signal is extracted for analysis.
- reinforcement learning** The study of how a biological organism or manmade intelligent agent learns to optimize its behavior based on its past experience with gaining rewards or other desired outcomes.
- resonant frequency** The frequency at which oscillations are most readily induced in a system, and at which it can oscillate at the greatest amplitude.
- retinal ganglion cells** The neurons in the retina that send visual signals to the thalamus.
- retinotopic organization**, or **retinotopy** The property of reinstating the spatial organization of the retina in the spatial organization of visual regions of the brain, such that, for example, adjacent regions of the retina are represented by adjacent regions of visual cortex.
- reuptake** The active process of removing molecules of neurotransmitter from the synaptic cleft by transporting them into a cell, either the presynaptic neuron or an adjacent astrocyte.
- saccade** A French word for a fast eye movement that “jumps” from point A to point B, with “no interest” in what is between the two points. Saccadic eye movements differ from smooth-pursuit eye movements, in which the eye follows the movement of an object in the environment.
- salience map** A neural representation of the location of stimuli that are currently the most relevant for behavior.
- scotoma** A small area of blindness in the visual field surrounded by normal visual sensation.
- seizure** An incidence of abnormal brain functioning that can last from seconds to minutes, and can occur as an isolated event or as a chronic condition of repeating seizures (i.e., epilepsy). *Positive* manifestations of seizure can include the perception of flashing light or the uncontrolled, involuntary jerking of a limb; *negative* manifestations can be a transient blindness or paralysis. Seizures result from stereotypical, abnormal patterns of activity in brain cells (neurons).
- semantic memory** Knowledge that was acquired through learning, but for which one cannot remember the context (e.g., place, situation, who was present) at the time that encoding of that information occurred.
- sensation** The detection of energy in the external world via one of the sensory organs (eyes, ears, skin, nose, tongue), or of the state of an internal organ via visceral sensors.
- sensorimotor circuits** Neural pathways connecting the processing of *sensory* information (i.e., having to do with vision, audition, touch, smell, or taste) with the *motor* system, which controls the movement of the body. The most basic sensorimotor circuits are reflex circuits that afford, for example, rapid withdrawal of the hand from a hot surface.
- signal detection theory** A framework in which decision-making factors are explicitly taken into account in tests of perception and recognition memory. Entails categorizing each response as a *hit*, *false alarm*, *miss*, or *correct rejection*.
- social cognitive neuroscience** One of many terms used to refer to the study of the neural bases of social interactions. (This field is the focus of *Chapter 16*.)
- somatotopy** This word derives from *soma* (Greek for *body*) and *-topy* (of or pertaining to a place). The principle of somatotopy is that adjacent parts of the body (for example, the upper lip and the nose) are represented in adjacent parts of the brain. Note that the “map” created by a somatotopically organized portion of cortex need not bear close

- physical resemblance to the parts of the body being represented – there are frequently distortions and discontinuities.
- spectrogram** A representation of how a power spectrum evolves over time.
- striate cortex** Another name for V1, so named for its striped appearance when viewed in just the right way.
- stroke** Disruption of the supply of blood to a region of the brain, either due to blockage (*occlusive stroke*) or bleeding from a vessel (*hemorrhagic stroke*).
- sulci** (*sing. sulcus*) Latin for “furrows,” fissures on the cortical surface.
- superior temporal polysensory area (STP)** The region in the superior temporal gyrus of the monkey that receives inputs from visual, auditory, and somatosensory regions, and in which many neurons respond to more than one sensory modality.
- synapse** Name introduced by Nobel Laureate Charles Sherrington (1857–1952) for the location at which one neuron passes a signal to another cell (within the CNS, it’s typically to the dendrite of another neuron; at the neuromuscular junction, it’s to a muscle cell). The name is derived from the Greek for “to clasp.”
- synaptic cleft** The gap between the presynaptic and postsynaptic terminals, across which neurotransmitter diffuses from pre- to post-.
- thermoreceptors** Receptors in the skin that are tuned to detect heat or cold.
- tonotopy** The organization of primary auditory cortex by frequency, which derives from the place code originating in the cochlear nucleus.
- topographic organization** The principle that adjacent parts of some aspect of the world outside the brain are represented in adjacent parts of the brain. Thus, somatotopy is one example of topographic organization. Other examples include the representation of the visual field, as projected onto the retina, by the visual cortex (“retinotopy”), the representation of sound in a continuous map from low to high frequencies (“tonotopy”), and the representation of geometric space (“grid cells” of the entorhinal cortex).
- transduction** The translation of physical energy from the environment into a neural code that represents the properties of that energy.
- traveling wave** A wave that propagates energy through a medium. In the case of the cochlea, it characterizes the propagation of vibrational energy from the oval window along the length of the basilar membrane.
- type vs. token** These two words are used to distinguish a category of stimulus (e.g., the category of faces) from a specific exemplar from within that category (e.g., my father’s face).
- typically developed (or “–ing”)** Refers to individuals who do not have a neurological or psychiatric diagnosis. In the context of studies of developmental disorders (e.g., autism spectrum disorder or fragile X syndrome), this term is often preferred to the synonym “normal,” to avoid the pejorative connotations of the label “abnormal.”
- utilization behavior** An exaggerated tendency for one’s behavior to be determined by the external environment.
- value-based decision** Selecting between two (or more) items when faced with a choice (as, for example, when at the store, or when deciding how to spend one’s Saturday evening). The purview of neuroeconomics; compare with *perceptual decision* and *foraging decision*.
- ventricle** A hollow chamber inside the brain through which cerebrospinal fluid (CSF) circulates. Like a chain of lakes connected by canals, the ventricles are part of a continuous system of CSF that extends to the caudal end of the spinal cord.
- vertical meridian** The imaginary vertical line that divides the visual field in half, thereby defining the boundary between the two hemifields.
- viral encephalitis** Acute inflammation of the brain due to viral infection.
- visual field** The full extent of what one can see while holding the eyes still, measured as degrees of angle from the point of fixation.
- visual hemifield** The half of the visual field represented in V1 of one hemisphere.
- voxel** A 3-D image element, the smallest unit of data in a PET or fMRI data set; 2-D “picture elements” having earned the nickname “pixel,” these “pixels with volume” have been dubbed “voxels.”
- white matter** Made up of billions of myelin-sheathed axons, the white matter makes up the majority of the volume of the cerebrum.

INDEX

Note: page numbers in *italic* refer to figures; page numbers in **bold** refer to boxes.

- Abrams, Daniel A., 518–519
- acoustic energy, 87, 88–89, 98
- action, 52, 53, 231
 - decision making, 424
 - oculomotor control, 231–254
 - semantic memory tasks, 344–349
 - skeletomotor control, 198–225
- action potentials, 34, 35–37
 - extracellular recording, **44**, 63, **64**
 - inhibitory surrounds, 72
 - long-term memory, 297
 - metabolic demands, **39**
 - motor learning, 217
 - oscillatory fluctuations, 43, 45–47
 - phototransduction, **57**
 - sensory receptor neurons, 99
 - thalamic neuron firing modes, 72, 73
 - V1 neuron as feature detectors, 63–65
 - vector summation, 211
- adenosine triphosphate (ATP), **39**
- Adlam, Anna-Lynne R., 352–354, 355
- adolescent decision making, **443**
- Adolphs, Ralph, 478, 479, 480, 485–487
- affective neuroscience, 478–491, 495
- age-related hearing loss, 98, **100**
- agnosia, visual, 264–266, 268, 275, 276, 344
- Aguirre, Geoffrey K., **239–241**
- Albus, James, 217
- Alexander, Garrett, 365–368
- allocentric localization, 124, 179–182, 302
- alpha-band oscillations, 74–77, **79–80**, 206, **380–381**, 547
- amnesia, 288–292, 322–325, 338, 362–363
- AMPA receptors, 37, 296–297, 375
- amygdala
 - emotion processing, 480–488, 495
 - fear extinction, 492, 493, **498**
 - memory, 303, **307**, 485–487, 492, 493
 - social behavior, **456–457**
 - trustworthiness judgments, 478–480
- anatomical specificity, 14, **15**
- Andersen, Richard A., 172–179
- anesthesia, 539, 540–543, 556–557
- animal (nonhuman) research, 11, 14–16, **17**, 18
- A-not-B error, 405–406, **407**, 408
- anterior cingulate cortex (ACC), 441–442, 453–455, 492, 493, 494, **498**, 538, 550
- anterior cingulate gyrus, 27
- anterior paracingulate cortex, 458
- anxiety disorders, **497–498**
- aperture problem, 149
- aphasias, 504–505, 506, 515, 516
 - primary progressive, **270**, 347–348, 349–356, **506**
- apperceptive agnosia, 264–265
- approach-related emotions, 491
- Armstrong, Kathleen, 243, **245**
- Assad, John A., 436, 437–438, 440
- association, memory, 291–300
- associative agnosia, 265–266, 268
- attention, 52, 53, 157, 231
 - consciousness and, **551**, 554
 - control of *see* attentional control
 - to locations vs. objects, 183–186
 - premotor theory, 221, 243
 - to sound, 93, 97, 183–184
 - spatial *see* spatial cognition and attention
 - to touch, 113
- attentional control, 229, 238
 - endogenous, 238–242, 243
 - exogenous, 238, 242, 243
 - via oculomotor system, 243–253
- sources, 238–242
- ToM and, 461, 464
- audition, 85–98, **100**, 183–184
 - see also* speech, perception
- auditory cortex (A1), 93–98, 99, 184–185, 509–512
- auditory evoked responses/potential (AEP)
 - attention, 183–184
 - auditory perception, 93, 95, 97
 - cognitive control, **403–404**, 407–408
 - dyslexia, **526**
 - speech production, 521–523

- auditory nerve, 88, 98
autism spectrum disorder (ASD), **225**,
464, **465–466**
autobiographical episodic memory, 306
avoidance learning, 484
axonal transport, 40, **41**
axon hillock, 34
axons
anatomy, 25–26, 27, 31
connectivity measures, 40, **41**, 42
diffusion tensor imaging, **32–33**
electrical and chemical processes, 34
myelination, 26, **27**
neuronal communication, 34
sensory receptor neurons, 99
- Baars, Bernard, 552
Baddeley, Alan, 368–369
Baddeley, Michelle, 467–473
Barbaro, Nicholas, 527
barrel cortex, 105–112
basal ganglia, 200, 201–202,
217–220
basilar membrane, 88–93, 98
Bates, Elizabeth, **524–525**
Bauby, Jean-Dominique, **536**
Baumgartner, G., 147–149
behavior, social *see* social behavior
Berger, Hans, 73–76
Berg, Esther, 401
Betz cells, 208
binding problem, 138
binocular rivalry, 546
biomechanics, 200–204
Bird, Helen, 516
Bisiach, Edoardo, 158, 159, 179
Blandin, Elise, **551**
Bliss, Timothy, 295, **296**
brain, 22
development, 24, **27**
evolution, 232, 301–303, 342,
489, 491
function *see* brain function
gross anatomy, 12, 24–30
orientation terms, 24, 25, 30
see also neurons, anatomy; neurons,
physiology
brain–computer interface (BCI),
211–213, **214**
brain function, 52–53
defining, **10**, 19
experimental perturbation, **130**,
280–283
localization, 8–19
mass action, 8–12, **13**, 142–143
topographic organization, 14
see also retinotopy; somatotopy;
tonotopy
see also action; attention; high-level
cognition; perception; sensation
brain lesion studies, 11, 13–17
long-term memory, 304, 339–342
short-term memory, 374–375
unilateral neglect, 155–160
visual object recognition, 124–132
voxel-based symptom lesion
mapping (VSLM), 157,
524–525
brainstem, **12**
auditory, 93, 94
consciousness, 534–535, **536**,
540, 543
brainstem oculomotor nuclei,
232–233
Brewer, James, 315–317
Britten, Kenneth H., 424–426
Broca, Paul, 16–17, 18
Broca's area, 18, **225**, 505, 506,
516–519, 523, **525**
Brown, Colin M., 520
Brown, Emery, 540
Buckner, Randy, 315, **316**, **350**
Bucy, Paul, 482
Burke, Christopher J., 467–473
Buszáki, György, 47, 302–303
- Cahill, Larry, 485–487
calcarine fissure, 33, 60
calcium (Ca⁺⁺) ions, 34, 297, 375
Caramazza, Alfonso, 339, 341–342
Cardin, Jessica A., **107**, **108–112**
Carlén, Marie, **107**, **108–112**
Carmena, Jose M., **212–213**
Casali, Adenauer, 556
Casey, B. J., **443**
category specificity
semantic memory, 339–342
ventral stream, 271–283
“caudal,” meaning, 25
cell-based neuroanatomy, 31, 40, **41**, 42
center-surround inhibition, 72
center-surround receptive field, 64
- central nervous system (CNS), 24
see also brain
cerebellum
gross anatomy, **12**, 24, 200
localization of function, 11
motor control, 200, 214–217
cerebral cortex
columnar organization, 30, 66–68,
143, **144–145**
see also barrel cortex
connectivity diagram, 133–134
connectivity measures, 40, **41**, 42
consciousness, **536**, 538, 539–544,
545–551, 552–557, 558
cortico-basal ganglia-thalamic
circuits, 200, 201–202, 218,
219, 430
gross anatomy, **12**, 24–27, 28–30,
40, 42
hippocampus memory role, 293,
300, 301, 303, **321**
localization of function, 11, 14–16, 18
oculomotor system organization,
232–233, 235, 237
plasticity, 112–116
synchrony control, 415–417, 418
see also auditory cortex (A1); motor
cortex; neurons, physiology;
somatosensory cortex; visual
cortex; *specific regions*
cerebrospinal fluid (CSF), 28
cerebrum, 24–27
see also cerebral cortex
Changeux, Jean-Pierre, 552, 553
Chao, Linda, **403–404**, 407–408
Chatham, Christopher H., 451–452
Chelazzi, Leonardo, 191–192, 549
chloride (Cl⁻) ions, **36**, 37–38
Chomsky, Noam, 504
Chun, Marvin, 272–275
clinical neuropsychology, **267**
cochlea, 87–93, 94, 98
cochlear implants, **96**, **98**
cognition, high-level *see* high-level
cognition
cognitive aging, 98, **100**
cognitive control, 397–418
adolescents, **443**
controller of, 408, 417–418
lateral frontal-lobe syndrome,
399–405, 407–408

- models, 405–413
 - neural activity, 413–417
- cognitive neuropsychology models, 266–268
- cognitive neuroscience, 2–3, 4
 - history, 6–7, 8–21
- cognitive psychology, 3
- Cohen, Jonathan D., 399
- Cohen, Neal, 305
- coincidence detection, 295–297, **299–300**
- color, semantic memory, 344–349
- color perception, 66, **132**
- coma, 534–535, 538–539, *544–545*
- competition, 158, 190–192
- complementary learning systems, 300, 301, 338
- computational models
 - attentional control, 246, **247–249**
 - cognitive control, 406–408
 - grammar, 515–516
 - memory, 300–301, 304
 - semantic memory, 340–341
 - social learning, 467–473
 - visual object recognition, 266–271
- computed tomography (CT), **161, 162**
- conditioning, 291, 293, **307**, 482–484, 492–493, **497–498**
- cone cells, **57**, 60–61
- conscientiousness, **10**
- consciousness, 2, 200, 306, 531–558
 - attention and, **551**
 - neural correlates, 533, 545–551
 - physiology, 533, 534–544
 - theories, 533–534, 552–557
- consolidation, 291, 306, **307–308**, 485
- construct validity, **10**
- contextual reinstatement, 326–328
- contralateral delay activity (CDA), **379–380**
- coordinate transformations, 172–179
- Corbetta, Maurizio, 241–242
- coronal plane, 25, 26, 30
- corpus callosum, 12, 26, 27
- cortical magnification, 61, 104
- cortical-reorganization hypothesis, 115–116
- cortico-basal ganglia-thalamic circuits, 200, *201–202*, 218, 219, 430
- corticocortical projections, 29, 70
- corticospinal tract, 198, 199
- corticothalamic loops, 28, 69–71, 72
- cotransport, **39**
- covert attention, 243, **244–245**, 246
- Craighero, Laila, 243
- Crick, Francis, 552
- cytochrome oxidase, 66, **68**
- Damasio, Antonio, 347, 478, *479*, 480
- Damasio, Hanna, 347
- Dapretto, Mirella, **225**
- Davidson, Richard, 491, 496
- Decety, Jean, 459–460, 464–466
- decision making, 422–446
 - action and, 424
 - adolescents, **443**
 - foraging, 424, 441–442
 - free will, **444–446**
 - perceptual, 424–433, 548
 - value-based, 424, 433–441, 451, **456**
- declarative long-term memory, 290, 291, 312–332, 485–487
- deep brain EEG, 543
- deep brain stimulation (DBS), **220**
- Dehaene, Stanislaus, 549–551, 552, 553
- Dehaene-Lambertz, Ghislaine, 509, **510–511**
- de la Vega, Alejandro, 451–452
- delayed gratification, 451–452
- delayed-response task, oculomotor, 237, 238, **239–241, 369–370**
- delay-task performance, 365–368, **369–370**, 371, 386–391
- dementia, 349
 - age-related hearing loss, **100**
 - frontotemporal, 458
 - semantic, **270**, 347–348, 349–356, **506**
- demyelinating diseases, **27**
- dendrites, **27**, 30, 31, 34, 35, 43
- 2-deoxyglucose imaging (2DG), **67, 69**
- depolarization, neurons, 34, 35, 43, 46–47
 - memory, 297, 301
- depression, **498**
- Desimone, Robert, 188–190, 191–192, 371–374, 375, 549
- D’Esposito, Mark, **239–241**, 400
- developmental cognitive neuroscience, 405–406, **407**
- Diamond, Adele, 405
- diffusion tensor imaging (DTI)
 - tractography, **32–33**
- distributed systems (mass action), 8–12, **13**, 142–143
- Ditterich, Jochen, **431–432**
- domain-specific knowledge, 339–340, 341–342
- dopamine, **37**, 218–221, 408, 409–413
- “dorsal,” meaning, 25, 30
- dorsal stream
 - language, 509–512, 516, *517*
 - visual processing, 160–179
 - consciousness, 554
 - vs. ventral stream, 124–135, 138, 160–162, **164–165**, 214, **215–216**
- dorsolateral prefrontal cortico-basal ganglia-thalamic circuit, 200, 202, 218
- dorsomedial frontal cortex (dmFC/DMFC), 233, 236, 413–414, 415, **416**, 441–442, **443**
- double dissociation of function, 124, **126, 343**, 516–517
- drift-diffusion models, **428–429, 431–432**
- drug mechanisms, **40**
- Duncan, John, 191–192, 517, *518*, 549
- Dungan, James, 462, **465–466**
- dyslexia, **526**
- ear, 87–88
 - see also* audition
- economic choice (neuroeconomics), 424, 433–441, 451, **456**
- effERENCE copy, 198
- egocentric localization, 124, 155–157, 169, 179, 182–183
- Eichenbaum, Howard, 305
- electrical properties of neurons, 31–34, 35–38, **39**, 43–47
 - see also* action potentials
- electrocorticography (ECoG), **44**, 280–283, 527
- electroencephalography (EEG)
 - auditory cortex (A1), 95
 - consciousness, 540–543, *544–545*, 556
 - motor cortex, 206–208
 - spatial attention, 186–188

- electroencephalography (EEG) (*cont'd*)
 spectral decomposition of data,
 91–92
 visual cortex, 73–81
see also event-related potentials (ERPs)
- electrophysiological experiments, **14, 15**
 electrodes, **44, 45, 66, 75, 281, 282**
 frontal-lobe syndrome, 402,
 403–404, 407–408
 fusiform face area, 280–283
 grammar, 519–520
 hippocampal memory circuits,
 293–295, **296**
 memory encoding, 315
 retrospective MVPA, **387**
 short-term memory, 365–368,
 369–375, **376–381, 386–391**
 somatotopy, 99–104
 spatial cognition, 172–179, 184–192
 temporal cortex, 135–138, 141–143,
 144–145, 280–283
 visual cortex, 63–68, 73–81
- emergent brain functions, **13**
- emotion, 476–498
 amygdala, **307, 478–488, 492, 493, 495, 498**
 extinction, 491–493, **497–498**
 memories, **307–308, 485–487, 492–493**
 mental health, 478, 491, 495,
 497–498
 range of perspectives on, 488–491
 regulation, 493–495, **496–498**
 trustworthiness, 478–480
- Emrich, Stephen, 384, **385**
- endogenous attentional control,
 238–242, 243
- Engel, Andreas, 430, **433**
- Engell, Andrew D., 478–480
- environmental dependency syndrome,
 400, 402–405
- epilepsy, 100–104, **281, 288**
- episodic memory, 300–303, 306
- equipotentiality, 11, 14
- Erickson, Cynthia A., 373–374, 375
- error monitoring, 413–414, **416**
- error negativity (Ne), 413, **416**
- error-related negativity (ERN), 413,
 414, 416
- ethics, **17**
- event-related potentials (ERPs)
 audition, 95, 97
 cognitive control, 402, **403–404, 407–408**
 language, 509, **510–511, 519–520**
 memory, 315, **379–380**
 spatial cognition, 183–188
- evolution, 232, 301–303, 342, **489, 491**
- excitatory neurons, **106, 108–109**
- exogenous attentional control, 238,
 242, 243
- explicit (declarative) long-term
 memory, 290, 291, 312–332,
 485–487
- express saccades, 235
- extinction, 158
 emotion control, 491–493, **497–498**
- extracellular electrophysiological
 experiments, **44, 45**
 orientation selectivity, 63, **64, 66**
 retrospective MVPA, **387**
 short-term/working memory,
 365–368, 369–374, 386–391
 spatial cognition, 172–179
 visual object recognition, **144–145**
- extrastriate cortex, 60, **132, 272–273**
see also MT (V5); V4
- eye, 56, 58, 59
 binocular rivalry, 546
see also visual sensation and
 perception
- eye movements, 231
 assessment tasks, 237
 attentional control, 229, 238–253
 conjugate, 231
 coordinate transformations, 172–179
 decision making and, 426–430, 435
 oculomotor system organization,
 232–237, 435
 saccadic category, 231–232
 smooth-pursuit, 231
 vergence shifts, 231
- face cells, 136, **137, 138, 143**
- faces
 recognition, 138–139, 143, **145, 266, 271–283**
 responses to fearful, 488, **489–490**
 trustworthiness judgments, 478–480
- Falkenstein, Michael, 413, **414**
- Fallah, Mazyar, **244–245**
- familiarity process, 328–332
- Farah, Martha, 264, 265–266, 268,
 340–341, 344
- fear conditioning, **307, 482–484, 492–493, 497–498**
- fear extinction, 492–493, **497–498**
- fearful face responses, 488, **489–490**
- feature-based attention *see* object-
 based attention
- feature detectors, 63–66, 69, 81,
 140–141
- Federmeier, Kara D., 520
- Fedorenko, Evelina, 517, **518**
- feedforward/feedback
 auditory transduction, 98
 consciousness, 547–549, 553–554
 visual cortex, 29, 69–73, 122–123,
 134, 143–149
- Feredoes, Eva, 462
- Ferrarelli, Fabio, 540–543
- Ferrier, David, 16, **18, 235–236, 453**
- Finger, Stanley, **11**
- Fisher, Simon, **514**
- Flourens, Pierre, 11, 14
- foraging decisions, 424, 441–442
- fornix, **12, 26**
- Fortin, Norbert J., 329–332
- forward model, speech, 521–523
- Fourier transform, **91, 94**
- foveal vision, 56, **57, 60–61**
- free will, **444–446**
- Fries, Pascal, 430, **433**
- Frith, Christopher D., 458–459
- Fritsch, Gustav, 14, **15**
- frontal cortex
 decision making, 436, **437–438, 440, 441–442, 443, 446**
 emotion extinction, 492–493
 frontal eye field, 233, 235–236, 237,
 241–242, 243–249, 250–253
 frontal-lobe syndrome, 399–405,
 407–408
 gross anatomy, **12, 24, 26**
 localization of function, 14, **15, 17, 18**
 performance monitoring, 413–414,
 415, 416
see also Broca's area; motor cortex;
 prefrontal cortex (PFC)
- frontotemporal dementia (FTD), 458
- Funahashi, Shintaro, **369–370, 371**
- function, brain *see* brain function

- functional connectivity analyses, **321**
- functional magnetic resonance
 imaging (fMRI), **39**, 162–169
 block-designs, **239–241**,
 272–273, 278
 computational, 470, **472**
 consciousness, 546–547, 549–551
 decision making, **444–446**
 emotion, 487–488
 event-related designs, 237, **239–241**,
 321
 face recognition, **272–273**, 276, 277,
 278–280, **282–283**
 fMRI adaptation, **350**
 language, **510–511**, 518–519
 memory encoding, 315–320, **321**
 memory retrieval, 326–327
 model-based, 470, **472**
 multivariate pattern analysis,
 251–252, 278–280, 326–327,
 382–386, **444–446**
 observational learning, 467–473
 oculomotor control, 237
 racial stereotyping, **481**
 semantic memory, **350**, 351
 short-term memory, 382–386
 spatial maps, 169–172, 303
 theory of mind, 459, 460–462,
 464–467
 trustworthiness judgments, 478–480
 univariate analyses, **274**, 384
 vegetative state, **537–538**
 visual attention, 186–188
- functional tissue imaging, 66,
 67–68, 69
- fusiform gyrus
 face area (FFA), 272–283
 semantic memory, 345–349
 unconscious vision, 549–551
- Fuster, Joaquin, 260, **270**, 365–368
- GABA (gamma-aminobutyric acid)
 GPI neurons, 218
 inhibitory effect, 37–38
 membrane-spanning channels, **36–37**
 neuroactive drug mechanisms, **40**
 optogenetics, **107**
 oscillating membrane potentials, 43,
 45–46
 reuptake, 38
- Gabrieli, John, 315–318
- Gaffan, David, 304
- Gage, Phineas, 452–453, 454
- Gallagher, Helen L., 458–459
- Gallese, Vittorio, **222–223**, **225**
- Gall, Franz Josef, 8, 11
- gamma-band oscillations, 43,
 107–112, **281**
- Gastaut, Henri, 206
- Gauthier, Isabel, 275–276
- Gehring, William, 413, 414
- genetic language disorders, 512–515
- Georgopoulos, Apostolos, 208–211
- Geschwind, Norman, 504–505
- glial cells, 25
- Glimcher, Paul W., 433, 434–435,
 440–441
- Global Workspace Theory, 552, 553,
 554, 555
- globus pallidus, internal segment
 (GPi), 218, 219, 220, 221
- glutamate
 NMDA receptors, 295–297,
 299–300, 301, 320
 postsynaptic effects, 37, 297
 reuptake, 38
 sound transduction, 88
- Gold, Joshua I., 427, **428**
- Goldman-Rakic, Patricia, 369–371,
 376, 405
- Goodale, Melvyn A., **215–216**
- Gopnik, Myrna, 513, **514**
- Gosseries, Olivia, 556
- grammar, 507, 512–520
- grandmother cells, 139, 141–143
- Granger causality analysis, 415–417
- Grant, David, 401
- Greebles, 276, 277
- gray matter, 25, 26, **32**
- Grill-Spector, Kalanit, **350**
- Gross, Charles G., 11–12, 18, 135–136,
 137, **245**
- Guillain-Barré syndrome, **27**
- gyrus, brain anatomy, 26, 27, 198
 see also fusiform gyrus
- Hagoort, Peter, 520
- hair cells, audition, 88, **96**, 98
- Hamker, Fred, 246, **247–249**
- hand cells, 135–136
- Hanks, Timothy D., **431–432**
- Hariri, Ahmad, 478
- Haxby, James, 160, **164–165**, 278–280,
 478–480
- Haynes, John-Dylan, **444**
- hearing, 85–98, **100**, 183–184
 see also speech, perception
- Hebbian plasticity, 291–295, **296**, 375
- Heeger, David, 546–548
- hemianopia, **60**
- hemispatial neglect, 155–160,
 169–172, 179, 242
- Hickok, Gregory, 509–512
- high-level cognition, 19, 396
 cognitive control, 397–418, **443**
 consciousness, 531–558
 decision making, 422–446
 emotion, 476–498
 language, 502–527
 motor system functions, 221, **222–225**
 social behavior, **443**, 449–473
- hippocampus
 anatomy, 24, 26, **293**
 memory, 182, 288, 291
 amnesia cases, 288, 289–290, **292**
 encoding, 300–301, 314–320, **321**
 episodic, 301–303
 long-term potentiation, 293–300
 map function, 302–303, 305,
 322–324
 retrieval, 328–332
 short-term, 305, 318–320
 standard model, 303–306
 spatial cognition, 179–182, 301,
 302–303, 322–324
- Hitch, Graham, 368–369
- Hitzig, Eduard, 14, **15**, **20**
- H. M. case, 288–292, 322–324, 338,
 342–343, 362–363
- Hodges, John, 352
- homunculus
 motor, 205–206
 sensory, 104, 105, 115–116
- horizontal plane, 25, 26
- Hoscheidt, Siobhan, 303, **323**
- Houde, John F., 521–522
- Hubel, David, 63–65, 66–67
- Humphreys, Glyn W., 266–268
- Hurst, Jane A., 512, 513, **514**
- hypercolumns, visual cortex, 66–68
- hyperpolarization, neurons, **34**, 37–38,
 43–47
- hypothesis generation, 155–160

- hypothesis testing
 - with fMRI, **274**
 - higher brain centers, 71, 547
- ibotenic acid injections, 304
- illusions, visual, 143–149
- imagery, 349
 - vegetative state patients, **537–538**
- imagined experiences, 305
- imitation behavior (IB), 400
- immunohistochemistry, **68**
- implicit bias processing, 480, **481**
- inferotemporal (IT) cortex
 - face recognition, 138–139, 143, **145**, 272–283
 - object-based attention, 191, 192
 - place representations, 182
 - short-term memory, 371–375, **387**
 - synchrony control, 415–418
 - visual object recognition, 122, 124–131, 133–136, 138–139, 141–146, 150
 - see also* face recognition *above*
- inhibitory interneurons, 37, **107**, **108–112**, 191
- inhibitory surrounds, 71, 72
- Integrated Information Theory (IIT), 555–557
- intention, 157–158, 169–170, **216**, 254–255
- interactivity, visual cortex, 69–81
- interneurons, **108**
 - GABA release, 37
 - object-based attention, 191
 - optogenetics, **107**, **108–112**
 - oscillating membrane potentials, 43–46
- intracellular recordings, neuronal, **44**
- intraparietal sulcus (IPS), 172–179, 237, 241–242, 250–253
 - see also* lateral intraparietal area (LIP)
- inverse problem, **74**
- ion channels, 34, 35–37, **38**, **39**, 43
 - memory, 297, 301, 375
- Irwin, Wil, 491
- isoluminance, 131
- Jacksonian march, 14
- Jackson, John Hughlings, 13–14
- Jacobsen, C. F., **364**
- Jerde, Trenton A., 250–253
- Kable, Joseph W., 433, 440–441
- Kanwisher, Nancy, 271–275, 458–459, 517, 518
- Kastner, Sabine, 415–417, 418
- Kiehl, Kent, **456**
- Klüver, Heinrich, 482
- Klüver–Bucy syndrome, 480–482
- Knight, Robert, 400, 402, **403–404**, 407–408, 527
- knowledge, 338
 - lack of visual *see* visual agnosia
 - object recognition model, **270**
 - semantic memory
 - acquisition, 338
 - category-specific deficits, 339–342
 - definitions, 338–339
 - domain-specific hypothesis, 339–340, 341–342
 - meaning, 338–339
 - neuroimaging, 342–349, 354–355
 - progressive loss, 349–356
- Koch, Christof, **551**, 552
- Koenigs, Michael, 455, **456**
- Koster-Hale, Jorie, 462, 464, **465–466**
- Kralik, Jerald D., **376**
- Kubota, Kisou, 365–368
- Kutas, Marta, 520
- Lai, Cecilia, **514**
- Lamm, Claus, 459–460, 464–466
- Lamme, Victor, 548–549, **551**, 553–555
- landmark discrimination task, 124, **125–126**, 127–129
- language, 502–527
 - genetics, 512–515
 - grammar, 507, 512–520
 - localization of speech, **15**, 16–18, 19, 517–519
 - semantic memory, 344, 347, 351–354, 356
 - speech perception, 89–91, **225**, 506–512, **524–525**, **526**
 - speech production, 89, 520–525
 - Wernicke–Lichtheim model, 504–506
 - written, **526**, 527
- “lateral,” meaning, 25, 26
- lateral frontal-lobe syndrome, 399–405, 407–408
- lateral geniculate nucleus (LGN), 58, 59, 60
- classes of RGCs, 69
- contribution to V1 feature detection, 63–66, 69
- feedforward/feedback, 29, 69–73, 122–123, 134, 149
- sensation–perception boundary, 81
- lateral intraparietal area (LIP), 426–430, 434–435
- lateral orbitofrontal cortico-basal ganglia-thalamic circuit, 200, 202, 218
- Laureys, Steven, 535, **536**, 538
- learning, 288
 - avoidance, 484
 - hippocampal mechanisms, 300–301
 - long-term potentiation, 295–297, **299–300**
 - memory retrieval, 322–324
 - motor, **211–212**, 217, 322–324
 - observational, 467–473
 - PDP model simulation, **269**
 - reconsolidation, **307–308**
- Lebedev, Mikhail A., **376**
- LeDoux, Joseph E., **307**, 483–484, 491
- lesion studies *see* brain lesion studies
- levodopa (L-dopa), 219–220, 221
- Lhermitte, François, 399, 400, 402
- Lichtheim, Ludwig, 505, 506
- ligand-gated ion channels, 35–37, 43
- light energy, **57**
- limbic cortico-basal ganglia-thalamic circuit, 200, 202, 218
- limbic lobe, **12**, 24, 27
- Lisman, John, 301, 302
- local field potential (LFP), 43
 - cortical synchrony control, 415–417, 418
 - EEG signal as, **74**
 - extracellular recording, **44**, 45
 - gamma-band oscillations, 43, **108–110**, **281**
 - motor cortex, 206–208
 - Parkinson’s disease, 220
- location-based attention, 183–190
- Lo, Chung-Chuan, **428–429**
- locked-in syndrome (LIS), 535, **536**, 539, 557
- locus coeruleus, 487
- Lomo, Terje, 295, **296**
- long-term memory (LTM) *see* memory, long-term (LTM)

- long-term potentiation (LTP),
293–297, **299–300**
- Luzzatti, Claudio, 158, 159, 179
- magnesium (Mg⁺⁺) ions, 297
- magnetic resonance imaging (MRI),
32, 33, 323
see also functional magnetic
resonance imaging (fMRI)
- magnocellular (M) neurons, 29, 69,
122, 123
visual processing stream, 131–132, 134
- Maguire, Eleanor, 305, 320–324
- Malmo, Robert, **364**
- Marr, David, 217
- Martin, Alex, 342–349
- mass action (distributed systems), 8–12,
13, 142–143
- Massimini, Marcello, **541–542**
- McCarthy, Rosaleen, 339, 344
- McClelland, James, 300, 301, 340–341,
344, 515–516
- McDermott, Josh, 272–275
- McNaughton, Bruce, 300, 301
- mechanoreceptors, 99, 105, 106
- “medial,” meaning, 25, 26
- medial temporal lobe (MTL)
emotion, 488
memory encoding, 315–318, 320, **325**
memory neurobiology, 288–289,
290–291, **292, 293,**
303–304, 305
memory retrieval, **325, 328–329**
semantic memory, 338
- meditation, **496–497**
- membrane potentials, 33–34, 43–47,
57, 74
see also action potentials
- memory, long-term (LTM), 286–309
amnesic patients, 288–291, **292,**
322–324, **325, 338**
association, 291–300
category-specific deficits, 339–342
complementary learning systems,
300, 301, 338
conditioning, 291, 293, **307,**
482–484, 492–493, **497–498**
consolidation, 291, 306,
307–308, 485
declarative, 290, 291, 312–332,
485–487
emotion, **307–308, 485–487,**
492–493
encoding, 300–301, 314–320, **321,**
325, 327–328
episodic, 300–303, 306
Hebbian plasticity, 291–295, **296**
long-term potentiation, 293–297,
299–300
multiple trace theory, 306
neuroimaging, 314–320, 326–327,
342–349, 354–355
NMDA receptors, 295–297,
299–300, 301, 320
nondeclarative, 290, 291
organizing principles, 290–291,
303–306
perception, 150, 260, **270, 304**
place cells, 182, 301, 302–303
plasticity, 288, 291–297, **299–300**
progressive loss, 349–356
reconsolidation, **307–308, 485**
relational binding, 305
retrieval, 301, 322–332
semantic, 300, 301, 306, 336–356
short-term memory and, 362,
368–369, **373**
spatial, 297, **299–300, 302–303, 305,**
320–324
standard model, 303–306
synaptic modification, 291–300
- memory, short-term (STM),
360–391
definitions, 362–363
hippocampus, 305, 318–320
incidental encoding, 318–320
long-term memory and, 362,
368–369, **373**
multivariate data analysis, 382–391
neuroimaging, 382–386
organizing principles, 291
prefrontal cortex, **364, 365–368,**
369–372, 374–375, **376–379,**
386–391
as prolonged perception, 362
working memory, 362, 363,
368–381, 388–391, 405–406
- meninges, 28
- mental health, 478, 491, 495, **497–498**
- mentalizing, 458–467
- mental lexicon, 338, 347, 507
- mental representation, 260
- memory *see* memory, short-term
(STM); memory, long-term
(LTM)
object recognition, 262–283
mental training, **496–497**
- Messinger, Adam, **376**
- Mesulam, M.-Marsel, 349–352, 354,
453–455
- Meyers, Ethan, **387**
- Michaelson, Laura, 451–452
- microstimulation
FEF, 236, **244–245**
LIP, **431–432**
MT, 424–426, **431–432**
PPC, 174–179
- Miller, Earl, 191–192, 371–374, 375,
376, 399, 549
- Milner, Brenda, 288–291, 322–324,
365, 401–402
- mind, theory of (ToM), 455–467
- mirror neurons, **222–225**
- Mitchell, Jason, 460–461, 464
- Molaison, Henry (H. M.), 288–292,
322–324, 338, 342–343,
362–363
- Molfese, Dennis L., **526**
- Monti, Martin M., **538**
- Moore, Tirin, 243, **244–245**
- moral judgments, 462–467
- Morris Water Maze, **299–300**
- Moscovitch, Morris, 306
- motion perception
aperture problem, 149
decision making, 424–433
feedforward pathway, 70–71
see also MT (V5)
- motor control *see* oculomotor control;
skeletal motor control
- motor cortex, 14, 198
anatomy, 198–200
cerebellar function, 217
cortico-basal ganglia circuits, 200,
201–202, 218, 219
decision making, **446**
dmFC performance monitoring, **416**
locked-in syndrome, **536**
motor homunculus, 205–206
neurophysiology, 206–214
perceptual decision making, 430–433
spinal cord motor computations,
204, 211

- motor evoked potentials (MEP),
206–207
- motor learning, **211–212**, 217,
322–324
- Motzkin, Julian C., **456–457**
- movement-related dynamics, 206–208
- MT (V5), 70–71, 122–123, 149, 171,
424–426, 430–433
- multiple sclerosis (MS), **27**
- multiple trace theory, 306
- multivariate pattern analysis (MVPA),
250–253, 278–280, 326–327,
382–391, **444–446**
- Mummery, Catherine J., 354–355, 356
- Munakata, Yuko, 406, 451–452
- mu rhythm, 206
- Murray, Elisabeth, 304
- muscles, 201–204, **231**
- myelin, 26
- myelination, **27**
- Naccache, Lionel, **551**
- Nadel, Lynn, 303, 306, **323**
- Nader, Karim, **307**
- Nagarajan, Srikanth S., 521–522
- navigational processing, 301–303
- near-infrared spectroscopy (NIRS), **509**
- neglect, 155–160, 169–172, 179, 242
- neocortex *see* cerebral cortex
- neostriatum, 200, 218–219
see also ventral striatum
- neural codes, 53, **57**
audition, 88, 98
oculomotor system, 233, 236–237
short-term memory, 386–391
skeletonotor control, 208–214
- neural correlates of consciousness
(NCC), 533, 545–551
- neuroactive drugs, **40**
- neuroeconomics, 424, 433–441,
451, **456**
- neuroimaging
block vs. event-related designs,
239–241, 314–315
long-term memory, 314–320, **321**,
323, 326–327, 342–349,
354–355
MVPA, 250–253, 278–280, 382–386
neuronal imbalances for, **39**
performance monitoring, **416**
physics, **32–33**
short-term memory, 382–386
theory of mind, 458, 459, 460–467
univariate analyses, **274**, 382, 384
visual object recognition, 271,
272–273, 276, 277, 278–280,
282–283
visual system, 160–172
see also functional magnetic
resonance imaging (fMRI)
- neuromodulatory tone, 218
- neurons, anatomy, 25–26, 27, 28–31
columnar organization, 30, 66
connectivity measures, 41
motor system, 198, 200
superior colliculus, 233
visual cortex, 28–30, 59–61
in vivo research, **32–33**
- neurons, physiology, 14, **15**, 31–47
attention, 183–192
attentional control, 243, **244–245**,
246, **247–249**, 250
cognitive control, 399
communication mechanisms, 34–38,
39, 45–47
connectivity measures, 40, **41**, 42
consciousness, 534, **535**, 540, **541**,
547, 548–549, 554
coordinate transformations, 172–179
drug effects, **40**
electrical/chemical properties,
31–34, 35–38, **39**, 43–47
see also action potentials
electroencephalography, 73–76
emotion extinction, 492–493
excitatory neurotransmission, 37
firing rate, **44**
hippocampal place fields, 179–182,
302–303
inhibitory neurotransmission, 37–38
mechanical energy transduction,
99, 106
memory, 291–300, 301–303, 304,
366–381, 386–391
metabolic demands, 38, **39**
mirror, **222–225**
oculomotor control, 232–237,
435, 554
optogenetics, **107**
pain transduction, 99
perceptual decision making, 424–430
phototransduction, **57**
population coding, 208–214
prefrontal cortex, 408, 409–413
skeletonotor control, 208–214,
216–217, 218–220, 221,
222–225
sound transduction, 88, 98
synchronous oscillation, 45–47,
108–112, 220, 221, 414–417
thermal energy transduction, 99
transcranial magnetic stimulation,
77, **79**
value-based decision making,
434–436, 438
vectorial hypothesis, 210–211
visual object recognition, 131,
135–149, 268–271
- neuropsychology, 11, **15**, 266–268
- neurosurgery, 100–104, **281**, 288
- neurotoxins, **40**
- neurotransmitters, 34, 35–38
consciousness, 534, 540
cotransport, **39**
drug mechanisms, **40**
excitatory, 37
see also glutamate
fear extinction, **498**
inhibitory, 37
see also GABA (gamma-
aminobutyric acid)
- memory, 295–297, **299–300**, 301,
320, 375
motor control, 218–220
oscillating membrane potentials, 43,
45–46, 220
PFC function, 408, 409–413
reuptake, 38, **39**
reward prediction error signal,
409–410
sound transduction, 88
- Newman, Joseph P., **456**
- Newsome, William, 424–426
- Nicolelis, Miguel, **211**, **214**
- nigrostriatal tract, 218–219
- Niki, Hiroaki, 365–368
- NMDA receptors, 295–297, **299–300**,
301, 320, **498**
- norepinephrine, 485
- nuclear magnetic resonance (NMR)
imaging, **32**, **33**
see also magnetic resonance imaging
(MRI)

- object-based attention, 158, 171, 183–188, 190–192, 243
- object recognition *see* visual object recognition
- observational learning, 467–473
- occipital cortex
 - gross anatomy, **12**, 24, 26–27
 - semantic memory, 354–355
 - speech perception, 512
 - visual function, **15**, 16, 18, 58–59, 60
 - see also* visual cortex
- Ochsner, Kevin, 493–494
- oculomotor control, 229–254
 - assessment tasks, 237, **239–241**
 - attentional control, 229, 238–253
 - consciousness, 554
 - decision making, 426–430, 435
 - locked-in syndrome, 536
 - PPC involvement, 174–179
 - system organization, 232–237, 435
- oculomotor cortico-basal ganglia-thalamic circuit, 200, 202, 218, 430
- oculomotor delayed-response (ODR) task, 237, 238, **239–241**, **369–370**
- O’Keefe, John, 179
- optical imaging, 67–68, **144–145**, **509**
- optic nerve, 59–60
- optic tract, 59–60
- optogenetics, **107**, **108–112**
- orbitofrontal cortex (OFC), 202, 436, 437–438, 440, 442
- orbitomedial PFC (omPFC), 435–440
- O’Reilly, Randall, 300, 301
- orientation columns, 66
- orientation selectivity, 63–66, 68–69
- orientation tuning, 65–66, 208–211, 213
- Ó Scalaidhe, Séamas P., **376**
- oscillatory synchrony, 45–47, **108–112**, 220, 221, 414–417
- Owen, Adrian M., **537–538**
- Padoa-Schioppa, Camillo, 436, 437–438, 440
- pain, 99, 115–116
- parallel distributed processing (PDP) models
 - cognitive control, **407**
 - grammar, 515–516
 - prefrontal cortex, 408, 409–413
 - semantic memory, 340–341
 - visual object recognition, 266, 267, 268–271
- parietal cortex, **12**, 24, 26–27
 - decision making, 426–430, 434–435
 - motor control, 214
 - oculomotor and attentional control, 237, 241–242, 243–246, 250–253, 435
 - place representations, 182
 - somatosensation *see* somatosensory cortex
 - spatial cognition, 172–179
 - visual object recognition, 122, 123, 124–131, 132
 - see also* temporoparietal junction (TPJ)
- Parkinson’s disease (PD), 200, 201, **206**, 219–220, 221
- Parvizi, Josef, 280–283
- parvocellular (P) neurons, 29, 69, 122, 123, 131–132, 134
- Pascual-Leone, Alvaro, 462–464
- Passingham, Richard E., 374, **514**
- Patterson, Karalyn, 352, 515–516
- Pavlovian conditioning, 291, 293, 482–484, 492–493, **497–498**
- Penfield, Wilder, 99–104, 205, 288
- perception, 52, 53
 - auditory, 93–98, 183–184, **403–404**, 407–408
 - see also* speech *below*
 - boundary with sensation, 81–82
 - expertise in, 275–276
 - memory and, 150, 260, **270**, 304
 - speech, 89–91, **225**, 506–512, **524–525**, **526**
 - STM as prolonged, 362
 - visual *see* visual object recognition; visual sensation and perception
- perceptual decision making, 424–433, 548
- performance monitoring, 413–414, **415**, **416**
- peripheral nervous system (PNS), **27**, 99, 106
- perseveration, 401–408, 412–413
- persistent vegetative state (UWS), 534, 535–539, 543–545, 556–557
- Perturbational Complexity Index (PCI), 556–557
- Peterhans, Ernst, 147–149
- Petersen, Steven E., 162–163, 342, 344
- Petrides, Michael, 374
- phantom limbs/pain, 115–116
- phase synchronization, neurons, 46, 47
- Phelps, Elizabeth, **481**
- phonemes, 506–509, **510**, **526**
- phosphenes, TMS-evoked, **79–80**
- photoreceptors, **57**, 60–61, **62**
- phototransduction, **57**
- phrenology, 8, 9, **10**, 11
- Piaget, Jean, 405
- Picard, Nathalie, 415, **416**
- Pillon, Bernard, 400, 402
- Pinker, Steven, 515, 516
- pinwheel model, 68
- place cells, 179–182, 301, 302–303
- place codes, audition, 89
- planes of section, 25, 26, 30
- plasticity
 - memory, 288, 291–297, **299–300**
 - somatosensory, 112–116
- Platt, Michael, 434–435, 439
- Plaut, David C., **343**
- Plum, Fred, **536**
- Poeppel, David, 509–512
- Poggio, Tomaso, **387**
- Pohl, Walter, 124, **125–126**, 127, **130**
- Polyn, Sean, 326–327
- population coding, 208–214
- positron emission tomography (PET), **39**, **161**
 - image subtraction, **161**, 162
 - memory encoding, 314
 - semantic memory, 342, 344–349, 354–355, 356
 - vegetative state (UWS), 538
 - visual system, 160–162, **164–165**
 - word processing, 162–163, 342
- Posner, Jerome B., **536**
- posterior parietal cortex (PPC)
 - coordinate transformations, 172–179
 - place representations, 182
 - visual object recognition, 122, 123, 124–131, 132
- post-traumatic stress disorder (PTSD), **497–498**

- potassium (K^+) ions
 Na^+/K^+ pump, **39**
 neuron physiology, 32, 33, **39**
 Powell, Lindsey J., 459
 prediction error signals, **409–410**, 469–473
 predictive coding
 language, 520
 motor control, 217
 visual system, 81–82
 prefrontal cortex (PFC)
 attentional control, 242
 cognitive control, 399, 402–414, 417–418, **443**
 consciousness, 538
 decision making, 435–440, 442, **443**
 developmental lag of myelination, **27**
 emotion, 491, 493, **498**
 fear extinction, 493, **498**
 memory encoding, 315–318, 320
 pyramidal neuron from, **31**
 semantic memory, 342–344
 short-term memory, **364**, 365–368, 369–372, 374–379, 386–391, 405–406
 social behavior, 451, 452–455, **456–457**, 458, 469, 470–473
 unique properties, 408–413
 see also Broca's area
 prefrontal cortico-basal ganglia-thalamic circuits, 200, 202, 218
 premotor cortex, semantic memory, 344–345
 premotor theory of attention, 221, 243
 priming tasks, 324, **325**, **350**
 propranolol, **307–308**, 485, 486, 487
 proprioception, 116–117, 172–179
 prosopagnosia, 266, 275, 276
 psycholinguistics, 521
 psychopathy, 455, **456–457**
 pulvinar nucleus, 415–417, **418**
 Purkinje cells, 216, 217
 pyramidal cells, 31
 glutamate release, 37
 identifying, **108**
 motor cortex, 198
 oscillating membrane potentials, 43–46
 place cells, 179
 prefrontal cortex, 408
 Quirk, Gregory, 492–493
 racial stereotyping, **481**
 Ramachandran, V. S., 115–116, **225**
 Ranganath, Charan, 318–320, **321**
 Rasmussen, Theodore, 205
 reading, dyslexia, **526**
 receiver operating characteristic (ROC) tests, 329–332
 receptive fields, 61, **62**, 65–66
 visual object recognition, **131**, 136–138, 140–141, 149
 recollection process, 328–332
 Recurrent Processing Theory, 553–555
 reentry hypothesis, attention, 246, **247–249**
 relational binding, 305
 repetition priming, 324, **325**, **350**, 351, 549
 response conflict, **416**
 Ress, David, 546–548
 reticular activating system, 534–535, 540, 543
 retina
 oculomotor system organization, 232–233, 235
 phototransduction, **57**
 processing information from, 63–81, 134, 232–233
 resolution, **58**
 visual system organization, **57**, 58–62
 retinal ganglion cells (RGCs), **57**, 59, **62**, 68–69
 retinotopy, 59–62, **69**
 spatial cognition, 169–170, 170–171, 172–174
 retrograde transport, **41**, 42
 reward prediction error signal, **409–410**, **469**, 470, **472**
 rhodopsin, **57**
 Ridderinkhof, K. Richard, 415, **416**
 Riddoch, M. Jane, 266–268
 Riggall, Adam, 384, 385
 Rizzolatti, Giacomo, 221, **222–223**, 243
 rod cells, **57**
 Rogers, Timothy T., **270**, **343**
 Romei, Vincenzo, **79–80**
 Romo, Ranulfo, **409**
 “rostral,” meaning, 25
 Rougier, Nicolas P., 408, 409–412
 Rushworth, Matthew, 374, 441–442
 Ryan, Lee R., 303, **323**
 Saalman, Yuri, 415–417, **418**
 saccades, 231–232
 attentional controller, 243, **244–245**, 246, **247–249**, 253
 oculomotor assessment task, 237, 238
 oculomotor system organization, 233, 234, 235, 236–237, 435
 sagittal plane, 25, 26
 salience maps, 241–242
 Salzman, C., 425, 426
 Saxe, Rebecca, 458–459, 460, 461–464, **465–466**
 Saygin, Ayse Pinar, 169
 Schacter, Daniel, 315, 349
 Schafe, Glenn E., **307**
 Schiller, Peter, 131–132
 Scholz, Jonathan, 461–462, 464
 Schultz, Wolfram, **409–410**, 410–412, 467–473
 Scoville, William, 288
 semantic dementia (primary progressive aphasia), **270**, 347–348, 349–356, **506**
 semantic memory, 300, 301, 306, 336–356
 acquisition, 338
 category-specific deficits, 339–342
 definitions, 338–339
 meaning, 338–339
 neuroimaging, 342–349, 354–355
 progressive loss of knowledge, 349–356
 single case studies, 342, **343**
 sensation, 52, 53
 auditory, 87–93, 98
 boundary with perception, 81–82
 somatosensation, 99–117, 172–179
 translation of information, 52–53, **57**, 88, 98, 99, **106**
 visual *see* visual sensation and perception
 sensory recruitment hypothesis, 382–386
 Serdaru, M., 400, 402
 Serences, John, 382–384
 Sereno, Martin I., 169
 Sergeant, Claire, 552, **553**
 Shadlen, Michael N., 427, **428**, **431–432**
 Shallice, Tim, 339, 340

- short-term memory (STM) *see*
 memory, short-term (STM)
- Shulman, Gordon, 241–242
- signal detection theory (SDT), **330**,
 427–428, **428–429**, **431**, 548
- signal processing, **91–92**, **93**, 94
- Sillito, Adam, 70–71, 71
- Simons, Jon, 349, 351
- skeletomotor control, 196–225
 biomechanics, 200–204
 cognitive functions, 221, **222–225**
 decision making, **446**
 localization, 13–14, **15**, 16, 18–19
 motor cortex anatomy, 198–200
 motor cortex neurophysiology,
 206–214
 motor homunculus, 205–206
 motor learning, **211–213**, 217,
 322–324
 motor system organization, 198–200
 outside of motor cortex, 214–221
 speech production, 520–521
- Slagter, Heleen, **496–497**
- sleep, 534, 539–540, **541–542**,
 556–557
- Smith, David V., 439–440
- social behavior, 449–473
 adolescent decision making, **443**
 observational learning, 467–473
 theory of mind, 455–467
 trustworthiness, 451–452, 473
 vmPFC, 452–455, **456–457**, 458,
 470–473
- social judgment processing,
 478–480, **481**
- sodium (Na⁺) ions
 Na⁺/K⁺ pump, **39**
 neuron physiology, 32, 33–34, 35,
39, 43
- somatosensation, 85, 99–117
 proprioception, 116–117, 172–179
- somatosensory cortex, 99, 101–102
 cell-based neuroanatomy, **41**, **42**
 phantom sensations, 115–116
 plasticity, 113–116
 somatotopy, 99–112
- somatotopy, 14, 99–112, 202, 205–206
- Somerville, Leah, **443**
- Soon, Chun Siong, **444–445**
- sound
 spectral decomposition, 88–93
 transduction, 88, 98
- spatial cognition and attention, 153–192
 allocentric localization, 179–182,
 302
 attention defined, 157
 competition, 158, 190–192
 control of attention, 238–253,
 461, 464
 coordinate transformations, 172–179
 dorsal visual processing stream,
 160–179
 egocentric localization, 124,
 155–157, 169, 179, 182–183
 extrapersonal (far) space, 158
 location-based attention, 183–190
 navigational processing, 301–303
 neuroimaging, 160–172, 186–188
 neurophysiology, 182–192
 object-based attention, 158, 171,
 183–188, 190–192, 243
 peripersonal (near) space, 158
 place, 179–182, 302–303
 premotor theory, 221, 243
 unilateral neglect, 155–160,
 169–172, 179, 242
- spatial long-term memory, 297,
299–300, 302–303, 305,
 320–324
- spectral analysis of time-varying
 signals, **91–92**, **93**, 94
- spectral decomposition, 88–93
- speech
 localization, **15**, 16–18, 19, 517–519
 perception, 89–91, **225**, 506–512,
524–525, **526**
 production, 89, 520–525
- Spekreijse, Henk, 548–549
- spike-field coherence data, 415–417, **418**
- spinal cord
 motor computations, 204, 211
 motor cortex anatomy, 198, **199**
 somatosensory pathways, 99, **101**, 115
- Spurzheim, Johann Caspar, 8
- Stedron, Jennifer Merva, 406–408
- Stern, Chantal, 314
- stimulus reduction studies, 141–143
- Stokes, Mark G., 386, 388–390
- striate cortex *see* visual cortex,
 primary (V1)
- Strick, Peter L., **415**, **416**
- stroke patients
 locked-in syndrome, **536**
 semantic memory, 339
 speech, 16–17, **18**, **524–525**
 unilateral neglect, 155–158, **159**, 179
 substantia nigra (SN), 200, **219–220**,
 233, 430
- sulci, **12**, 26, 60
 intraparietal, 172–179, 237,
 241–242, 250–253
- Supèr, Hans, 548–549
- superior colliculus (SC), 232, 233, **234**
- superior intraparietal sulcus (IPS),
 172–179
- superior temporal polysensory area
 (STP), 136, **137**, 138–139
- supplementary eye field (SEF), 233,
 236–237
- Sylvian fissure, **12**, 26, 30
- synapses
 action of drugs at, **40**
 development in prefrontal cortex, **27**
 memory, 291–300, 301, 303, 375
 metabolic demand, **39**
 neuronal communication, 34, 35,
 38, 297
 oscillating membrane potentials,
 43–45
- synaptic cleft, 34, 35, 38
- synchronous oscillation, 45–47,
108–112, 220, 221, 414–417
- synesthesia, **348**
- syntax *see* grammar
- systems neuroscience, **15**
- Tanaka, Keiji, 141–143, **144–145**
- Tanifuji, Manabu, 143, **144–145**
- Tarr, Michael, 275–276
- temporal cortex
 face recognition, 138–139, 143, **145**,
 272–283
 gross anatomy, **12**, 24, 26–27
 language, 509–512, 516
 memory–perception relation,
 150, 304
 object-based attention, 191, **192**
 place representations, 182, 301
 semantic memory, 344–349,
 353–355
 short-term/working memory,
 371–374
 synchrony control, 415–418

- temporal cortex (*cont'd*)
 unconscious vision, 549–551
 visual object recognition, 124–131,
 133–139, 141–146, 150
see also face recognition *above*
see also fusiform gyrus; medial
 temporal lobe (MTL)
- temporoparietal junction (TPJ), 242,
 458–467
- Teuber, Hans-Lukas, 265, 275, 324
- thalamic reticular nucleus (TRN), 71, 72
- thalamus
 burst mode neurons, 72–73
 consciousness, 534, 538, 540,
 543–544, 552, 556–557
 cortical synchrony control,
 415–417, 418
 cortico-basal ganglia-thalamic
 circuits, 200, 201–202, 218,
 219, **220**, 430
 gross anatomy, **12**, 24, 26, 28, 29
 short-term memory, 365–368
 somatotopy, 106
 thalamocortical projections, 29, 71,
 72, 73, 106
 tonic mode neurons, 72–73
see also lateral geniculate nucleus
 (LGN)
- theory of mind (ToM), 455–467
- thermal energy transduction, 99
- Thier, Peter, 174–179
- threat vigilance, 488, **489–490**
- time-varying signals, **91–92**
- Tobler, Philippe N., 467–473
- Todorov, Alexander, 478–480
- Tononi, Giulio, **541–542**, 555
- tonotopy, 96–98
- touch *see* somatosensation
- Tranel, Daniel, 478, 479, 480
- transcranial magnetic stimulation
 (TMS), **77–79**
 consciousness, 540–543, 556, 557
 language, 521
 social behavior, 461, 462–464
- trustworthiness, 451–452, 473,
 478–480
- Tsuchiya, Naotsugu, **551**
- Tulving, Endel, 314, 322
- Ullman, Michael T., 515, 516
- uncinate fasciculus, 457, 482
- unilateral neglect, 155–160, 169–172,
 179, 242
- unresponsive wakefulness syndrome
 (UWS, vegetative state), 534,
 535–539, 543–545, 556–557
- utilization behavior, 400
- V1 *see* visual cortex, primary (V1)
- V2, 122–123, 133–134, 140, 147–149,
 190
- V4, 140, 148, 149, 190, 243, 246,
247–249, 415–418
- V5 *see* MT (V5)
- value-based decision making, 424,
 433–441, 451, **456**
- van Berkum, Jos J. A., 520
- Vargha-Khadem, Faraneh, **514**
- vegetative state (UWS), 534, 535–539,
 543–545, 556–557
- Velly, Lionel, 543
- “ventral,” meaning, 25
- ventral stream
 language, 509–512, 516, 517
 visual processing
 category specificity, 271–283
 consciousness, 554
 vs. dorsal stream, 124–135, 138,
 160–162, **164–165**, 214,
215–216
 feedback, 143–149
 stimulus representation hierarchy,
 139–143
 temporal cortex neurons,
 135–139, 141–143
- ventral striatum, 469–471, 470–473, **498**
- ventricles, 28
- ventromedial prefrontal cortex
 (vmPFC)
 decision making, 440, 442
 fear extinction, 493, **498**
 social behavior, 452–455, **456–457**,
 458, 470–473
- vergence shifts, 231
- vibrations, auditory sensation, 89–93
- Vigand, Phillippe, **536**
- visual agnosia, 264–266, 268, 275,
 276, 344
- visual angle, 61
- visual attention *see* spatial cognition
 and attention
- visual cortex
 attentional control, 243, 246,
247–249, 250–253
 connectivity diagram, 133–134
 gross anatomy, 28–30, 33
 perceptual decision making,
 424–426, 430–433
 primary (V1), 28–30
 category representation, 278–280
 consciousness, 546–549, 552,
 553–554
 electroencephalography, 73–81
 feedforward/feedback, 29, 69–73,
 122–123, 134, 149, 547–549,
 553–554
 history of cognitive neuroscience, 16
 information processing, 63–81,
 134, 232–233
 interactivity, 69–81
 IT neuron properties and,
 136–138
 oculomotor system organization,
 232–233, 235
 predictive coding, 81–82
 retinotopy, 58–62, 69, 171
 sensation–perception boundary, 81
 sensory recruitment hypothesis,
 382–386
 stimulus representation hierarchy,
 139–141
 what vs. where (ventral vs. dorsal
 stream), 124–135, 165
 spatial cognition, 170, 171, 190
 synchrony control, 415–417, 418
see also MT (V5); V2; V4
- visual-evoked potentials (VEP),
 184–188
- visual field, 56–58, 59–60, 61, 231
- visual illusions, 143–149
- visual object recognition, 120–151,
 262–283
 amygdala, 478, 480
 faces, 138–139, 143, **145**, 266,
 271–283
 hierarchical nature, 139–143
 neuropsychology models, 266–268
 object- vs. image-based, **146**
 PDP models, 266, 267, 268–271
 perception–memory boundary, 150,
 260, **270**
 perceptive expertise, 275–276
 projections from V1, 122–123

- semantic memory, 340–341
- temporal cortex neurons, 135–139, 141–143
- ventral stream category specificity, 271–283
- ventral stream feedback, 143–149
- viewpoint dependence, 143, **146**
- visual agnosia, 264–266, 268, 275, 276
- what vs. where (ventral vs. dorsal stream), 124–135, 138, 160–162, **164–165**
- visual–proprioceptive information integration, 172–179
- visual sensation and perception, 54–82
 - action and, **215–216**
 - amygdala, 478, 480
 - cognitive neuroscience history, 14–16, 18–19
 - consciousness, 546–551, 553–554
 - coordinate transformations, 172–179
 - cortical synchrony, 415–417, *418*
 - expertise, 275–276
 - object recognition *see* visual object recognition
 - oculomotor system and, 232–233
 - sensation–perception boundary, 81–82
 - V1 information processing, 63–81
 - visual system organization, 56–62
 - visual working memory, **369–371**, 386
 - visuoperceptual decision making, 424–433
 - voltage-gated ion channels, 34
 - voltage-sensitive dye, **67**
 - Von der Heydt, Rüdiger, 147–149
 - voxel-based lesion–symptom mapping (VLSM), *157*, **524–525**
 - voxel-based morphometry (VBM), **323**, 354
 - Wagner, Anthony, 315–317
 - Wang, Gang, 143, **144–145**
 - Wang, Xiao-Jing, **428–429**
 - Warrington, Elizabeth, **325**, 339, 340, 344
 - Wearing, Clive, **292**
 - Weiskrantz, Lawrence, 11–12, **325**
 - Wernicke, Karl, **269**, 504–505
 - Wernicke–Lichtheim model, 504–506
 - Whalen, Paul, 478, **489–490**
 - what vs. how motor control, 214, **215–216**
 - what vs. where visual system, 124–135, 138
 - PET mapping, 160–162, **164–165**
 - visual working memory, **369–371**
 - whiskers, somatosensation, 106–112
 - white matter, 25, 26, **32**
 - whole-cell staining, *31*, **41**
 - Wiesel, Torsten, 63–65, 66
 - Wilson, Fraser A. W., **376**
 - Wisconsin Card-Sorting Task (WCST), 401–402, 412
 - Wisconsin General Testing Apparatus (WGTA), **125**, 127, 130, 401
 - Wise, Steven P., **376**
 - withdrawal-related emotions, 491
 - Wong-Riley, Margaret, 66, **68**
 - words, 507
 - dyslexia, **526**
 - processing, *162–163*, 342–344
 - semantic memory
 - aphasia, 351–356
 - category-specific deficits, 339–341
 - mental lexicon, 338, 347
 - neuroimaging, 342–349
 - working memory, 362, 363, 368–381, 388–391
 - A-not-B error, 405–406
 - X-ray computed tomography, **161**, **162**
 - Young, Liane, 462–464, **465–466**
 - Zarahn, Eric, **239–241**

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.